

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

Bradley Broom and John Weinstein,

Michael C Ryan, Chris Wakefield, Robert Brown, Futa Ikeda, Mark Stucky, David W. Kane, Rehan Akbani, James Melott  
Department of Bioinformatics and Computational Biology, UT MD Anderson Cancer Center

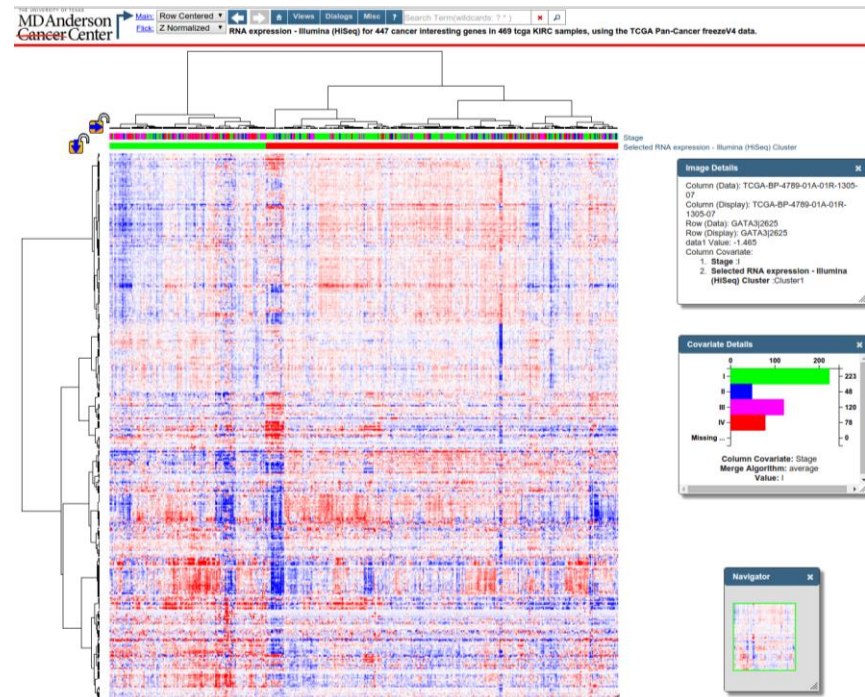
A visualization tool for dynamic  
exploration of large, omic data sets.

## Static Heat Maps

- Introduced by Weinstein et al. 1992
- Most frequent way of representing large omic data sets

## NG-CHMs: Visualization Features

- Zoom from summary views of very large data sets down to smallest details
- Fluid navigation
- Generate publication-quality graphics
- Highly interactive
- Scalable



# 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, cBioPortal
- Or to dynamic analyses
  - E.g. Box Plot, Points Plot
- 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
- Five major types of resource links

Row Label Menu
×

Selected Rows: 20

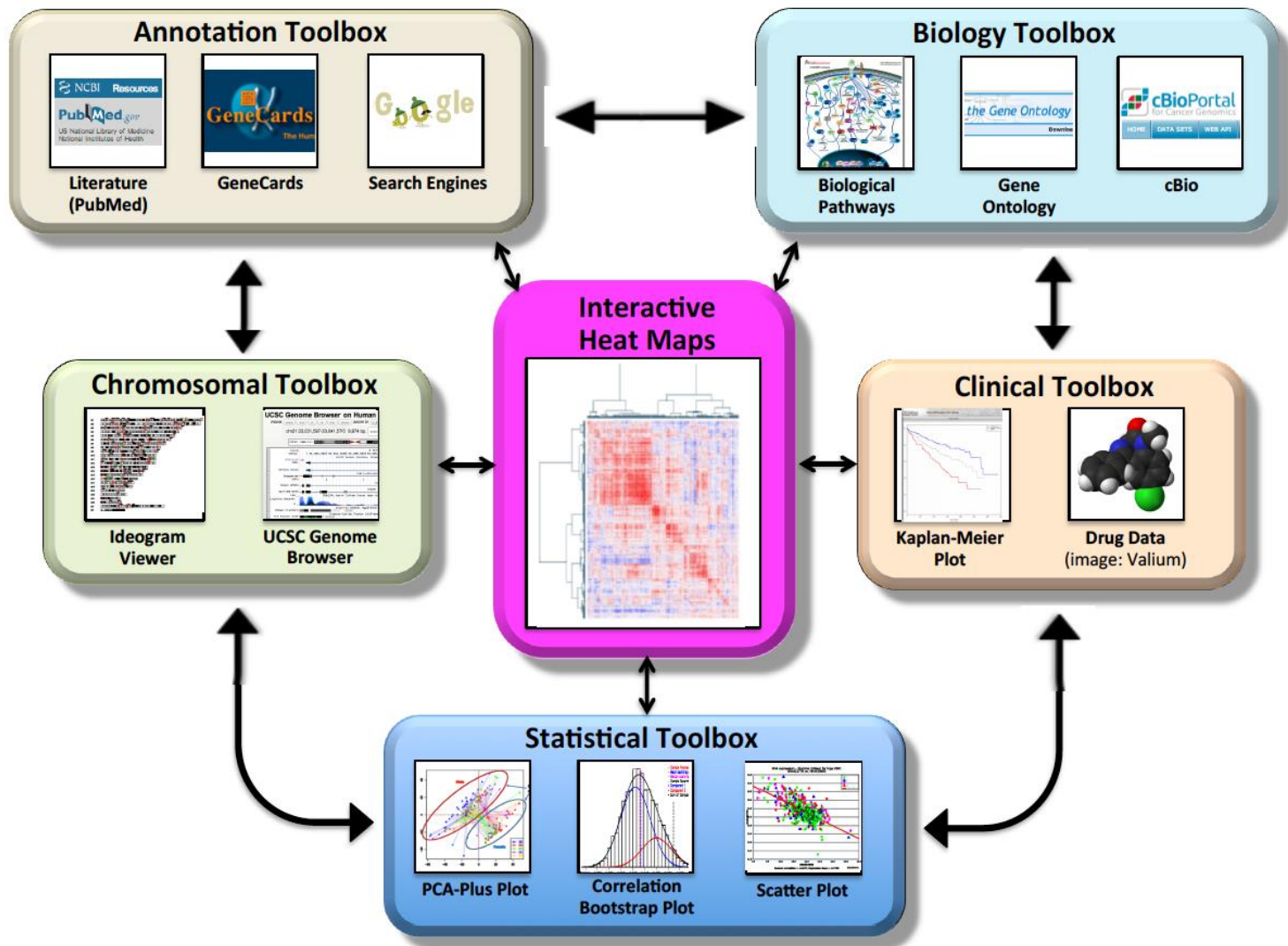
Copy to Clipboard  
View Box Plot  
View Points Plot  
View Genecard  
Search ClinicalTrials.gov for all  
View NCBI Gene  
View NCBI Entrez ID  
View Ideogram  
Search in Google Scholar  
Search Pubmed for all  
Search Pubmed for any  
View in cBio Portal

Column Label Menu
×

Selected Columns: 1

Copy to Clipboard  
View SlideArchive  
View in cBioPortal

# NG-CHMs in an Exploratory Environment



# Creating NG-CHMs

## Four “builder” levels:

### – Simple GUI

- New non-quantitative user

### – Advanced GUI

- Experienced non-quantitative user

### – R package

- 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	
Main Data Layer	Choose File	No file chosen

[Add](#) [Clear](#) [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 >>](#)

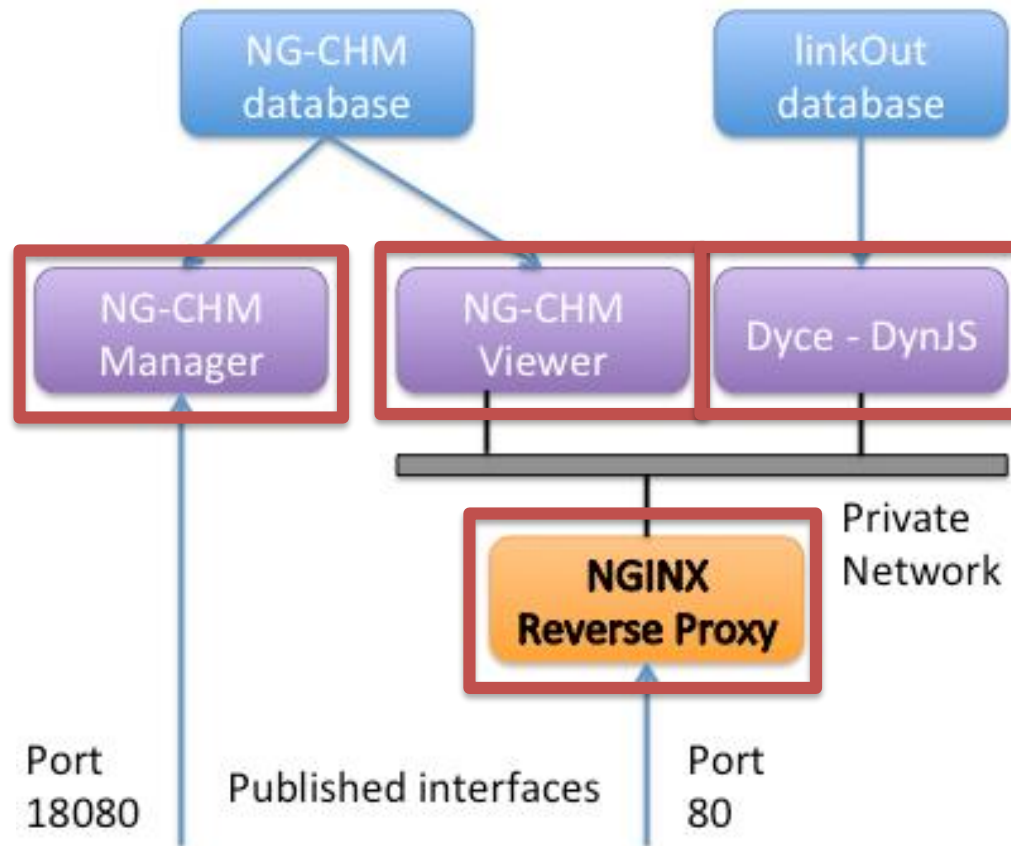
[Create Heatmap!](#) [Home](#)

---

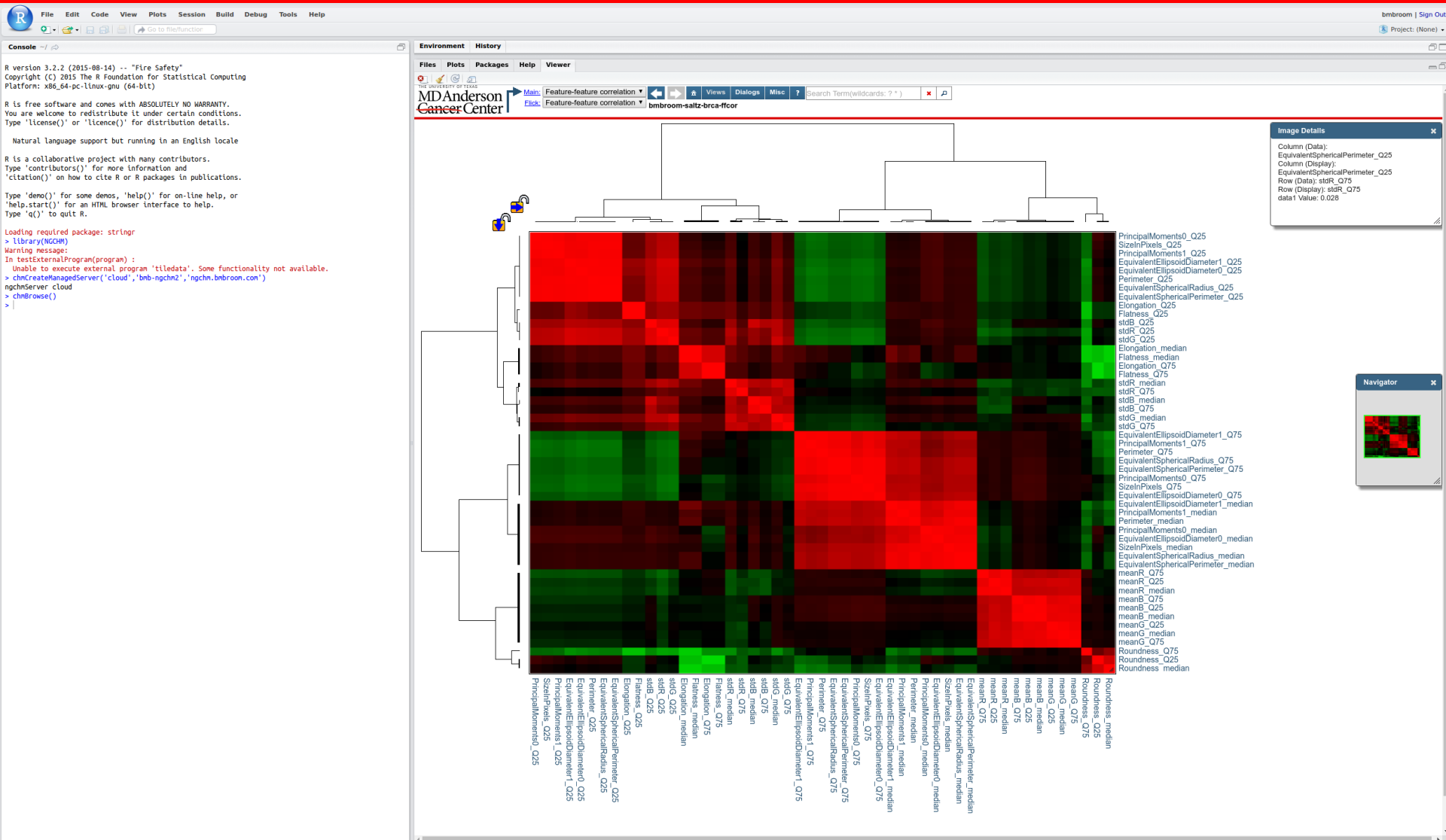
# RECENT DEVELOPMENTS

# 'Docker-compose' NG-CHM Container System

- Simplifies deployment of a full NG-CHM system to your own server or the cloud
- Includes support containers in addition to NG-CHM viewer:
  - NG-CHM Manager
  - Dynamic computation engine
  - Reverse proxy



# 'Docker-compose' RStudio container with NG-CHM





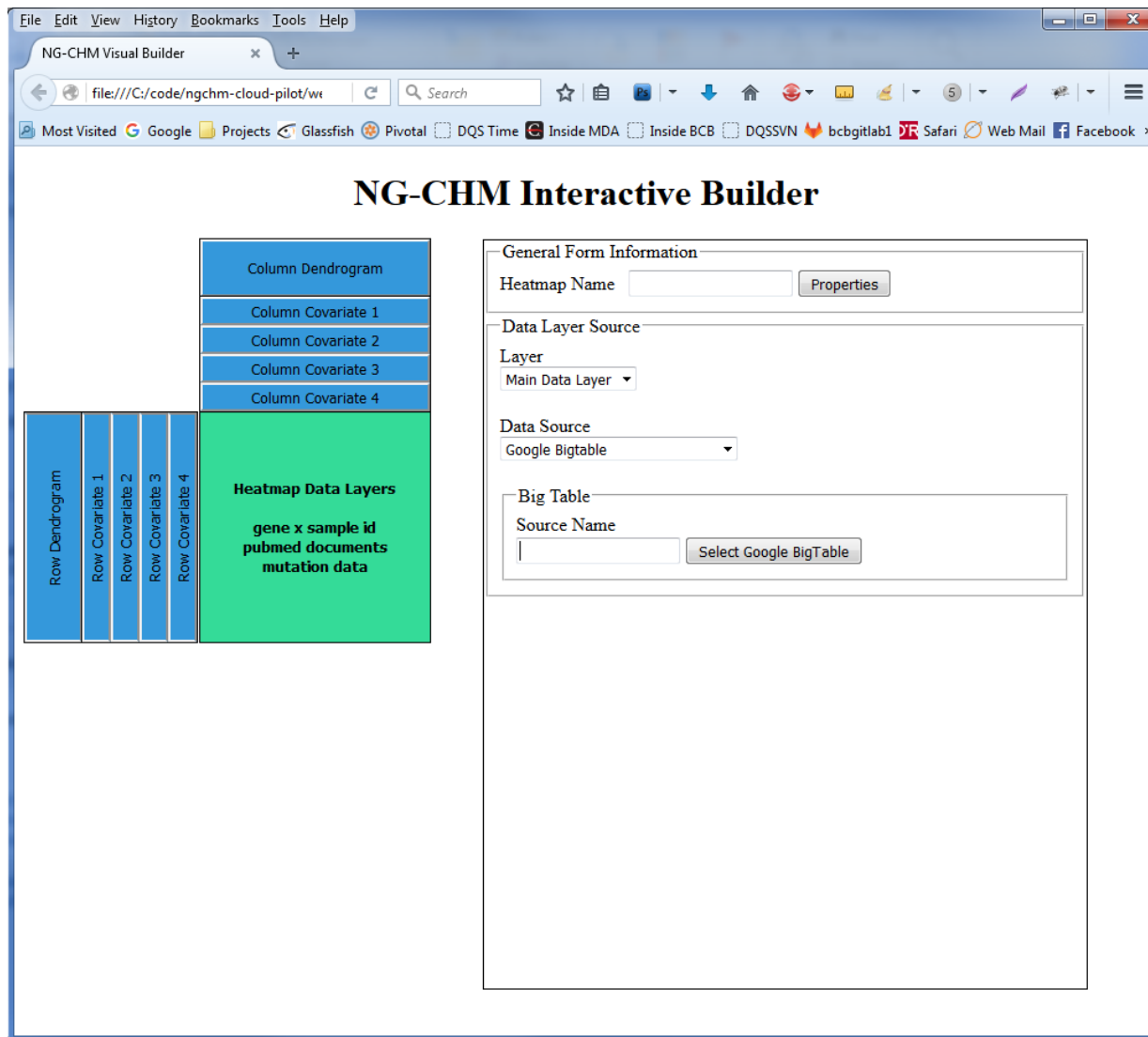
## NG-CHM tools for the ISB-CGC pilot

---

- Developed R library components for downloading data from Google BigQuery and the ISB-CGC pilot
  - Matrix data, clinical covariates, mutation status
- Created technical specification for an NG-CHM containing cloud and/or private data
- Commenced development of a graphical NG-CHM builder



# Graphical builder interface for NG-CHM ISB-CGC



The screenshot shows a web browser window titled "NG-CHM Visual Builder" with the address bar displaying "file:///C:/code/ngchm-cloud-pilot/wt". The browser's toolbar includes various icons for navigation and search. Below the browser window, the application interface is titled "NG-CHM Interactive Builder".

The interface is divided into two main sections. On the left, there is a vertical stack of components: a "Row Dendrogram" and four "Row Covariate" buttons (1, 2, 3, 4). To the right of these is a "Column Dendrogram" and four "Column Covariate" buttons (1, 2, 3, 4). Below these components is a large green box labeled "Heatmap Data Layers" containing the text "gene x sample id", "pubmed documents", and "mutation data".

On the right side of the interface, there is a "General Form Information" panel. This panel contains several input fields and buttons:
 

- Heatmap Name:** An input field with a "Properties" button next to it.
- Data Layer Source:** A section containing a "Layer" dropdown menu set to "Main Data Layer".
- Data Source:** A dropdown menu set to "Google Bigtable".
- Big Table:** A section containing a "Source Name" input field and a "Select Google BigTable" button.

## Transition to a new NG-CHM architecture

---

- Existing NG-CHM system architecture
  - Was designed more than five years ago
  - Relies heavily on back-end server for tile rendering
  - Does not exploit recent advances in web browsers
- New NG-CHM system architecture
  - Complete re-design to take advantage of new browser capabilities (e.g. WebGL)
  - Rendering performed entirely in browser
  - Improved performance and scaling characteristics
  - Far more responsive user experience

## New architecture NG-CHMs

---

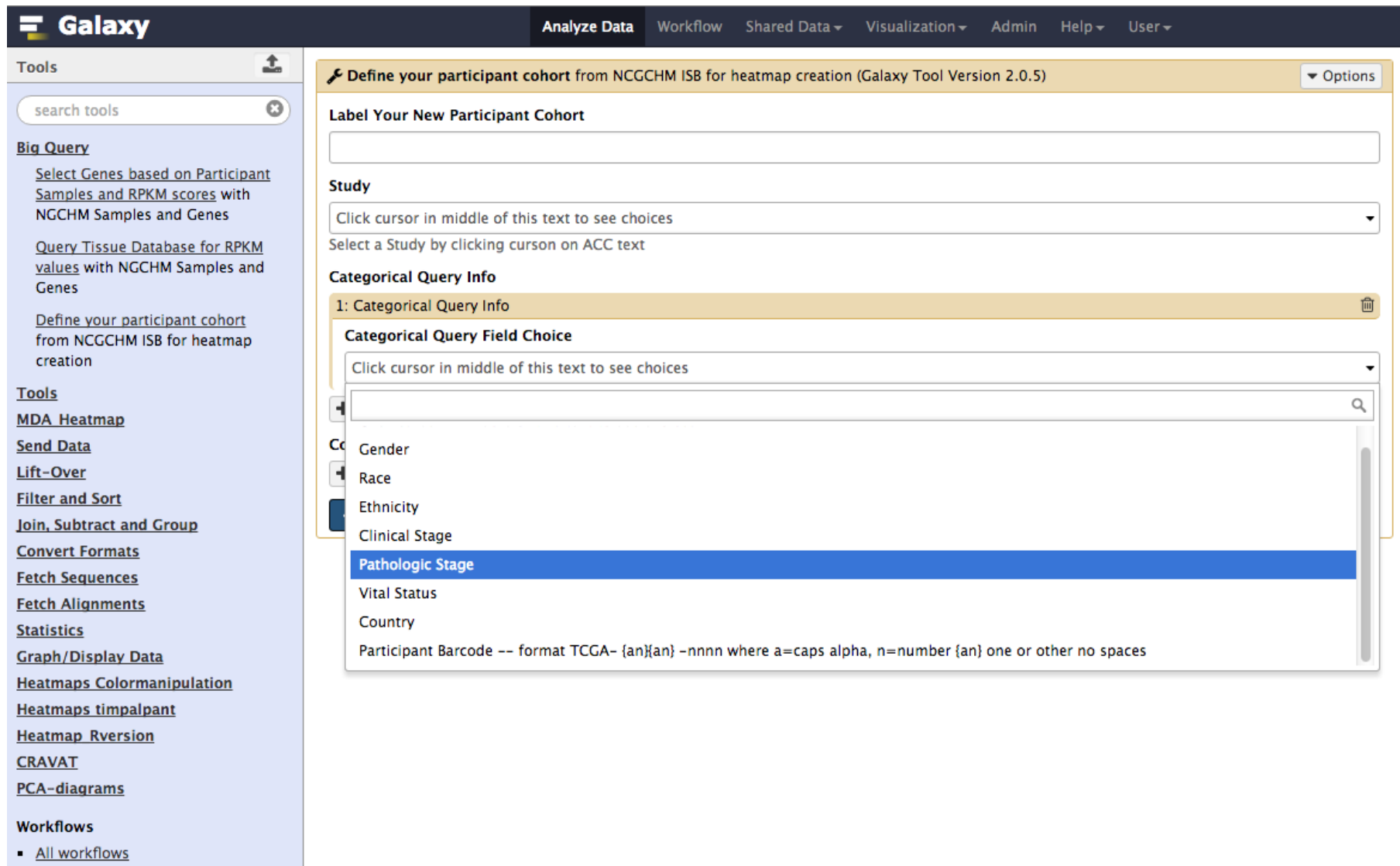
- Standalone server-less mode possible
  - Allows NG-CHMs to be:
    - Saved locally (e.g. with other project data)
    - Viewed while not connected to a network
    - Attached as supplementary material to publications
    - Shared with colleagues via email, web sites, portals, etc.
  - Nearly-all NG-CHM functionality is present
    - But cannot invoke server-side dynamic computations

## NG-CHMs integrated into Galaxy

---

- Added new architecture NG-CHMs to Galaxy:
  - BigQuery Interface (for ISB-CGC data access)
  - NG-CHM Builder tool
  - NG-CHM Visualization
- Exposes NG-CHMs to a new user community

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



The screenshot displays the Galaxy web interface for the 'Define your participant cohort from NCGCHM ISB for heatmap creation' tool. The interface is divided into a left sidebar with navigation links and a main workspace for tool configuration.

**Left Sidebar (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

**Main Workspace (Tool Configuration):**

**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 cursor on ACC text

**Categorical Query Info**

1: Categorical Query Info

**Categorical Query Field Choice**


Click cursor in middle of this text to see choices

**Field Selection List:**


- Gender
- Race
- Ethnicity
- Clinical Stage
- Pathologic Stage** (Selected)
- 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




[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

Unnamed history




2 shown

60.55 KB

2: Mean-Center a Matrix on data 1

1: matrix.tsv

# Galaxy NG-CHM Visualization Tool

Galaxy

Tools

search tools

Big Query

Tools

MDA Heatmap

Send Data

Link-Over

Filter and Sort

Join, Subtract and Group

Convert Formats

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

Heatmaps Colormanipulation

Heatmaps timpalpent

Heatmap Rversion

CRAVAT

PCA-diagrams

Workflows

All workflows

Analyze Data

Workflow

Shared Data

Visualization

Admin

Help

User

Using 417.5 MB

History

search datasets

Demo Define Participants

6 shown, 9 deleted

4.87 MB

15: HM\_name

295 lines

format: ngchm, database: ?

HM\_name\_

/home/rbrown/HM\_galaxy/databa

Hierarchical euclidean ward

Hierarchical euclidean ward

average

/home/rbrown/HM\_galaxy/tools/MI

/home/rbrown/HM\_galaxy/databa

/home/rbrown/HM\_gal

binary file

14: Normalize Matrix

on data 13

13: Mean-Center a Ma

trix on data 6

6: Query Tissue Datab

ase for RPKM values o

n data 5 and data 1

5: blca

167 lines

format: tabular, database: ?

QueryString = SELECT

ParticipantBarcode FROM [lsb-

cgctcga\_201510\_alpha.Clinical\_dat

WHERE ( Study = "BLCA" AND

gender= "MALE" AND race=

"WHITE" AND weight>=50.0 AND

weight<= 100.0 );

TCGA-E7-A3Y1

TCGA-FD-A43U

TCGA-FD-A43X

TCGA-DK-AA6T

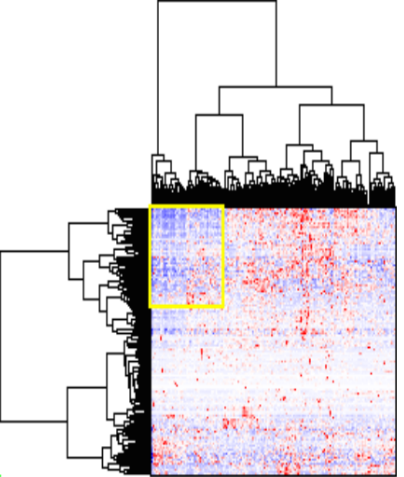
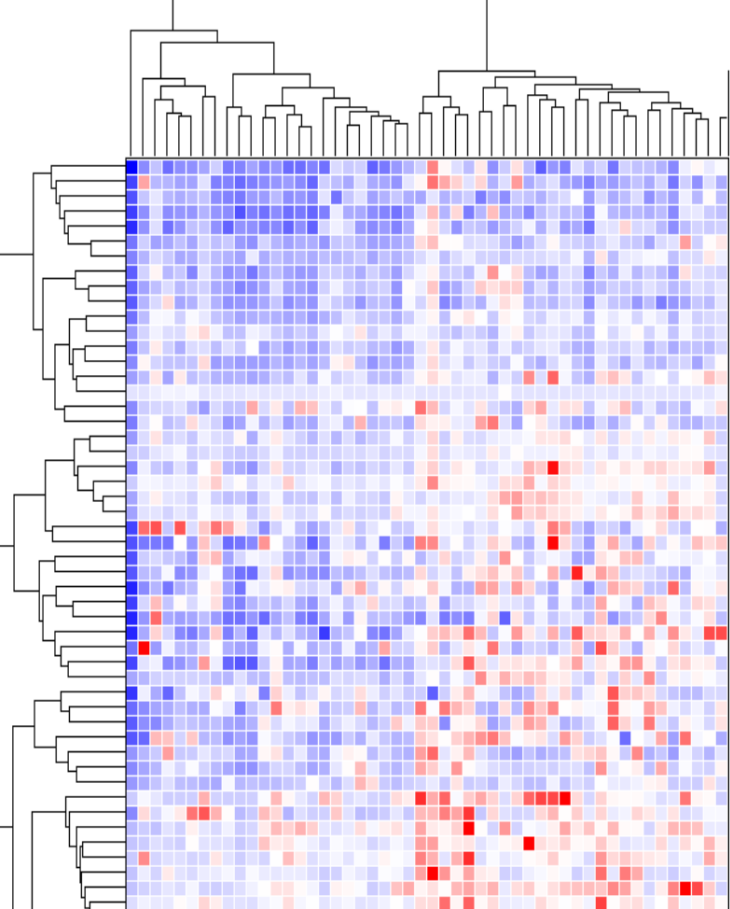
TCGA-DK-AA6R

1: oncogenes tsq tca

a.txt

Map Name: HM\_name

Search: GO

CREBBP  
 TET2  
 NF1  
 EP300  
 ARID1B  
 APC  
 ATRX  
 ARID1A  
 SETD2  
 PBRM1  
 SF3B1  
 MDM4  
 ARID2  
 BCOR  
 KDM6A  
 AR  
 RUNX1  
 ACVR1B  
 BRCA1  
 BRCA2  
 DNMT1  
 EZH2  
 MSH2  
 MSH6  
 SMARCA4  
 GNA11  
 TSC1  
 CBL  
 CIC  
 KDM5C  
 MED12  
 NRAS  
 CTNNB1  
 FUBP1  
 SKP2  
 PTEN  
 CYLD  
 JAK1  
 CASP8  
 RB1  
 NFE2L2  
 NCOA3  
 SOX9  
 NPM1  
 B2M  
 JAK2  
 PRDM1  
 NOTCH1  
 MET  
 MYC



---

# PLANS

## Plans

---

- Release new-architecture NG-CHMs
  - Starting with a beta release of the Galaxy tool
- Update our TCGA NG-CHM compendium to new-architecture NG-CHMs
- Scale-up NG-CHMs to larger data matrices
- Develop a tablet-optimized user interface
- Collaborate
  - Links to/from NG-CHMs from/to other (ITCR) tools
  - More advanced tools & data exchange

---

# VIDEO

<https://www.youtube.com/watch?v=v5G-HKIXI5U>