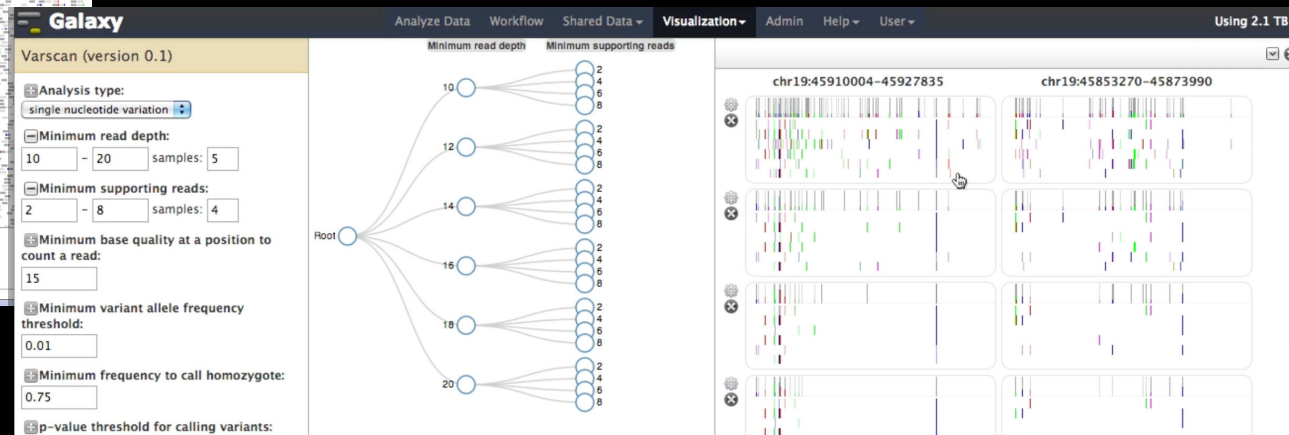
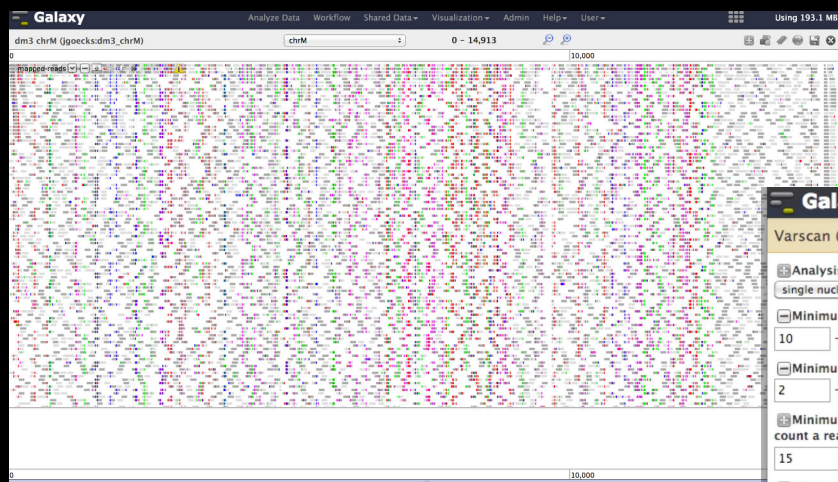




Galaxy Community Conference 2015: Visualization Workshop

Aysam Guerler and
Carl Eberhard

Jeremy Goecks

Recommended Web Browsers

Chrome will probably work best

Updated Safari/Firefox should work well

Internet Explorer and old versions of
Safari/Firefox may have problems

Topics

Visualization history and introduction

Numerical Visualizations

Biological Visualizations

Adding your own visualizations

Why Visualize?

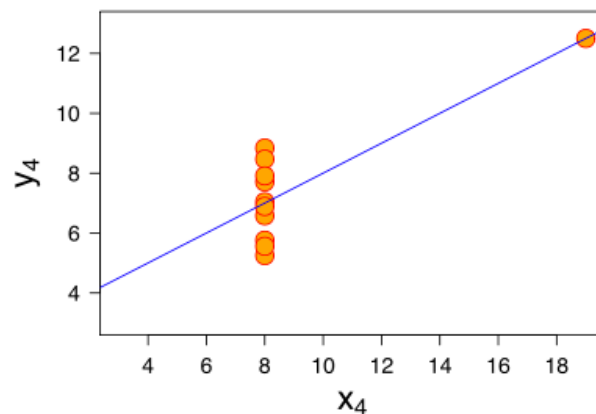
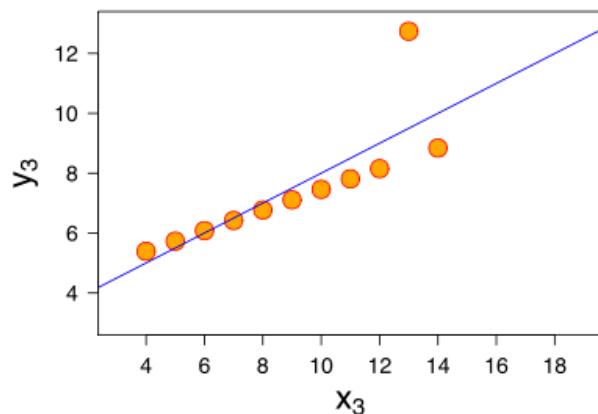
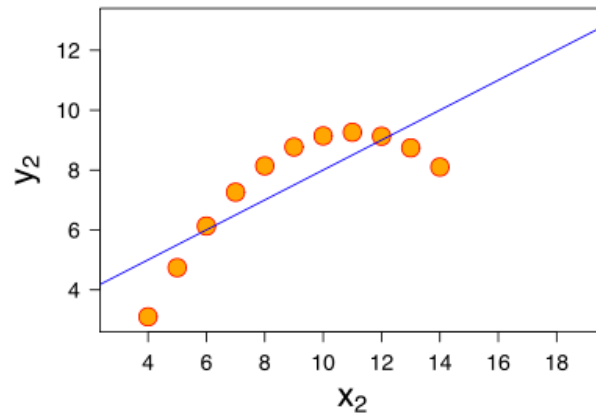
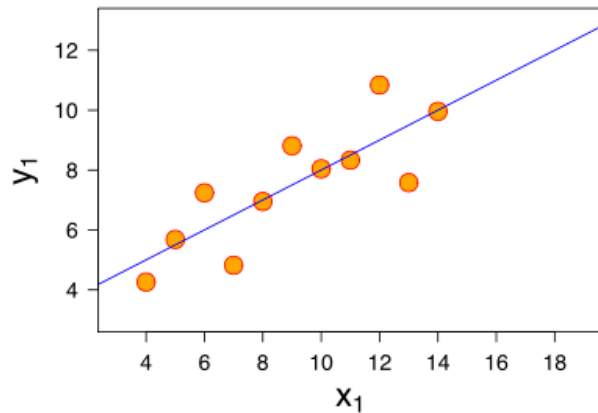
Why Visualize?

Quick check: did it work?

Exploration and hypothesis generation

Sharing/publishing

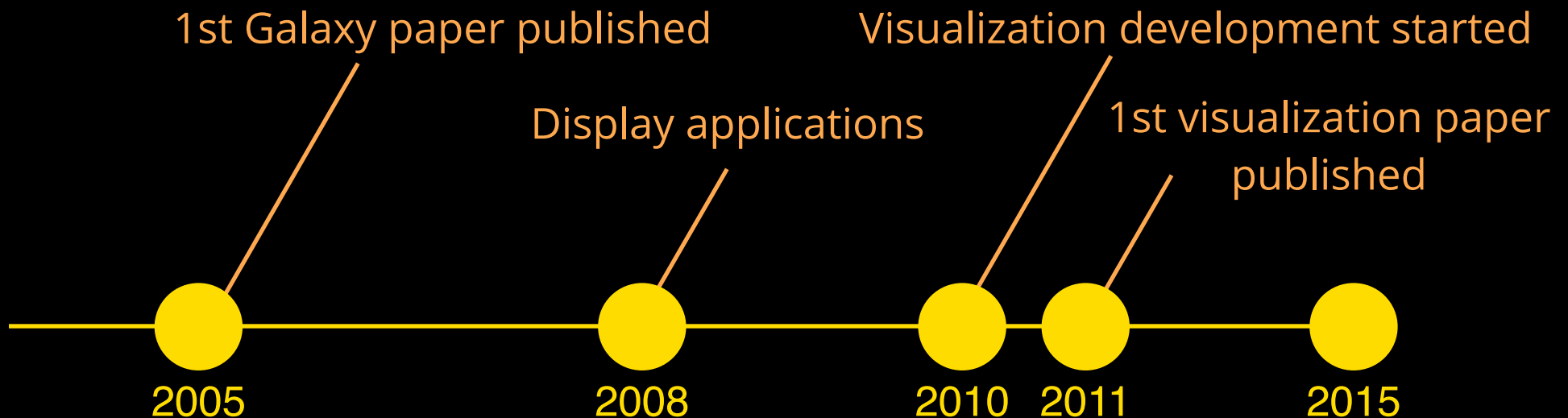
Anscombe's Quartet



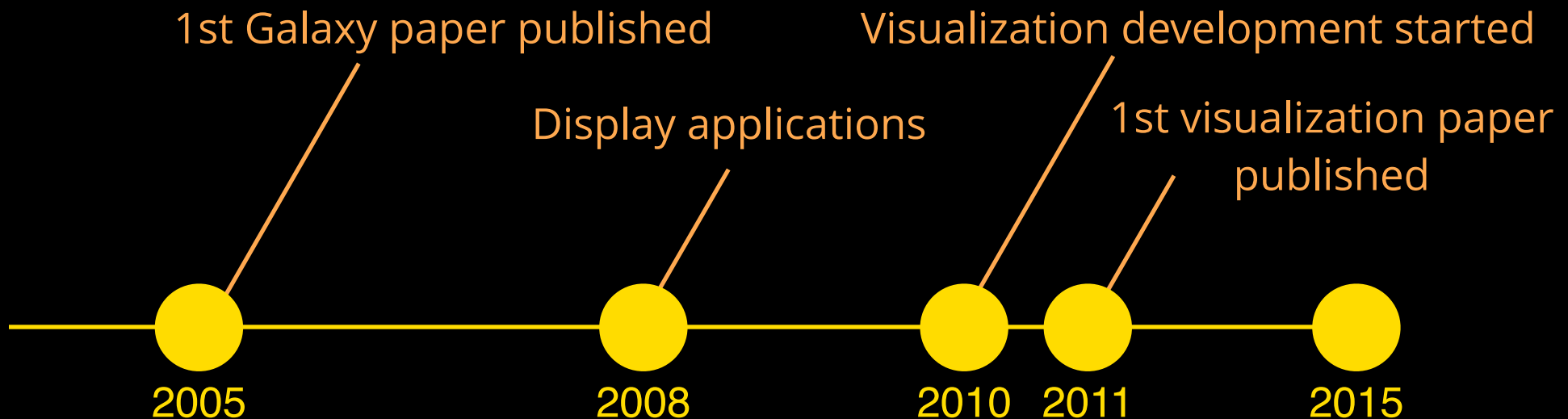
Property	Value
Mean x	9
Variance x	11
Mean y	7.5
Variance y	~4.125
Correlation	0.816
Linear regression	$y = 0.5x + 3$

http://en.wikipedia.org/wiki/Anscombe's_quartet

Timeline of Visualization in Galaxy



Timeline of Visualization in Galaxy



- 1. *visualization in Galaxy is nascent***
- 2. *you will be working with awesome new features***
- 3. *there may be bugs — help us fix them!***

Workshop Goals

Participants: learn about how to visualize your data in Galaxy

- ✦ biological visualizations
- ✦ numerical visualizations
- ✦ what Galaxy is doing underneath the covers

Instructors: feedback from you about what you like, don't like, and where to go next

Galaxy Visualizations

Visualizations are first-class objects in Galaxy, just like tools

A visualization can be added to Galaxy via a configuration file that specifies:

- ✦ datasets that can be used
- ✦ location of visualization code (client-side or on server)

Galaxy handles visualization integration and data management

- ✦ users can focus on analyzing data
- ✦ developers can focus on creating visualizations

Visualizations are 1st class Galaxy objects

Can be saved and versioned for reproducibility

Have a human-readable URL for sharing a fully interactive visualization:

`http://usegalaxy.org/u/jgoecks/v/tumor-mutations`

Can embed interactive visualizations in online supplementary materials via Galaxy Pages

Visualization Architecture

Client-server architecture



Lots of moving pieces

- ✦ prepare/process data on server
- ✦ send to client
- ✦ render on client

Topics

Visualization history and introduction

Numerical Visualizations

Biological Visualizations

Adding your own visualizations

Analysis goal: what similarities and differences can be found in cancer cell lines using exome and transcriptome sequencing?

Sequencing and Analysis

Sequenced exomes and transcriptomes of 3 pancreatic cancer cell lines

- ✦ MiaPaCa2, HPAC, and PANC-1

Datasets available in published history:

- ✦ Exome subset: *KRAS*, *STK11*, *ERBB2* aligned reads, removed dups, created read pileup
- ✦ transcriptome subset: *KRAS*, *STK11*, *ERBB2* aligned reads
- ✦ gene fusions from all cell lines
- ✦ whole transcriptome aligned reads coverage
- ✦ (gene annotation)

Display Applications

1: MiaPaCa2 Exome (KRA
S, STK11, ERBB2)

297.5 MB

format: bam, database: hg19

uploaded bam file





display at UCSC [main](#)

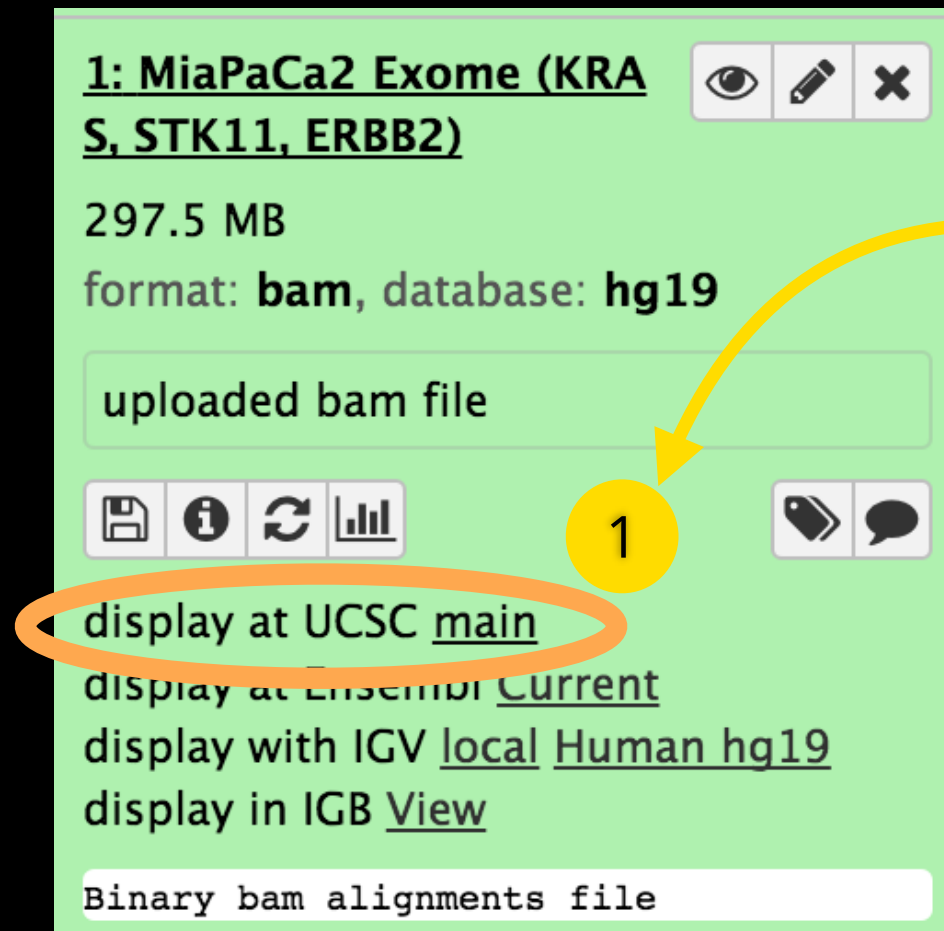
display at Ensembl [Current](#)

display with IGV [local](#) [Human hg19](#)

display in IGB [View](#)

Binary bam alignments file

Display Applications



1: MiaPaCa2 Exome (KRA S, STK11, ERBB2)   

297.5 MB
format: bam, database: hg19

uploaded bam file

    **1**  

display at UCSC main

display at Ensembl Current

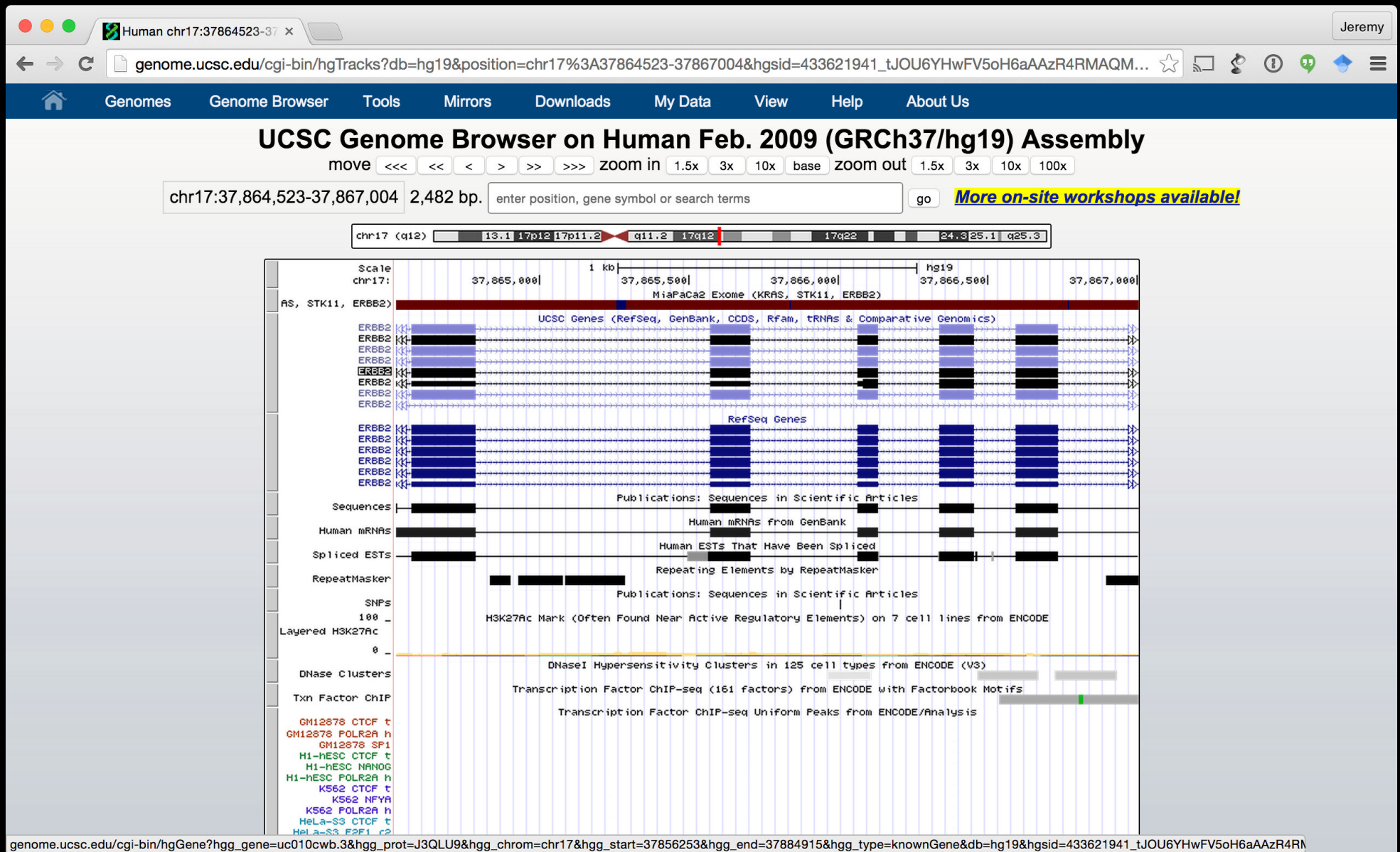
display with IGV local Human hg19

display in IGB View

Binary bam alignments file

Used throughout slides to show actions to take

Display Applications



Display Applications

Advantages

- ✦ use familiar tools
- ✦ easy to view your data alongside public datasets

Disadvantages

- ✦ cannot easily share/version visualization
- ✦ many more visualizations than display applications in Galaxy
- ✦ no data processing or visual analysis, only visualization

Trackster—Galaxy's Genome Browser



Trackster—Galaxy's Genome Browser

Genome browsers are a foundational genome visualization tool

Trackster is for the high-throughput sequencing era

- ✦ very large datasets, numerous simultaneous tracks
- ✦ maximum flexibility for customization (e.g. rainbow tracks)
- ✦ 2-3 indices per dataset for fast visualization

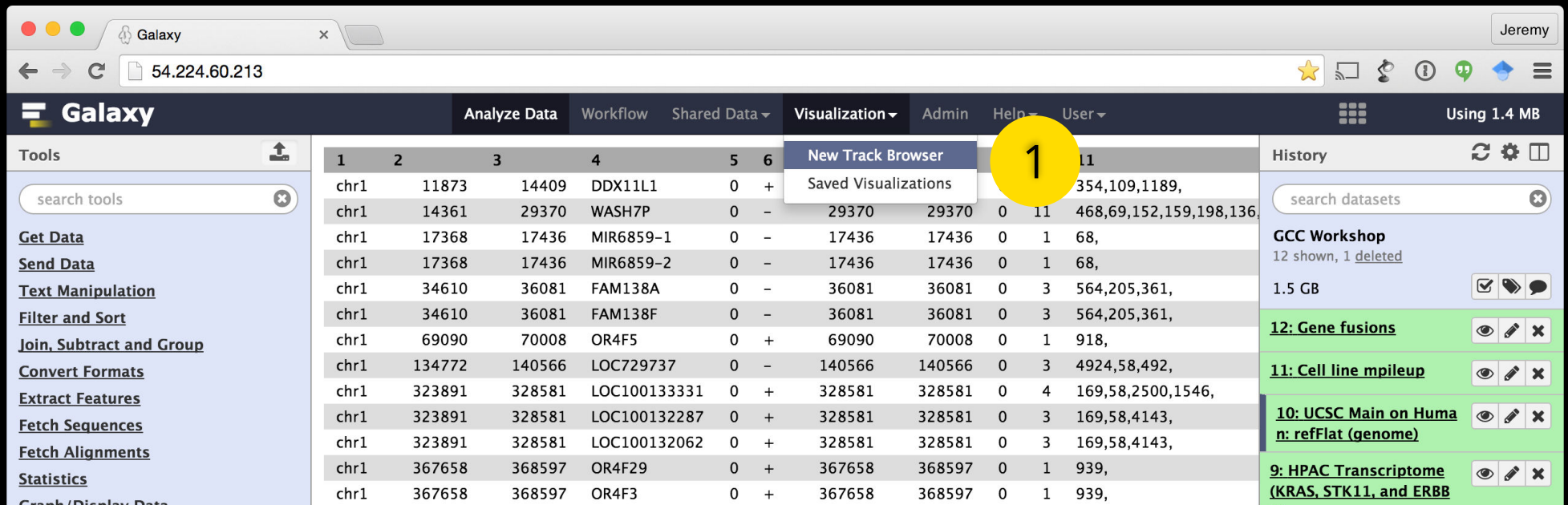
BED, GFF/GTF, interval, SAM/BAM, VCF, Wiggle, BigWig, BigBed, BedGraph

Let's visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)
3. Save visualization
4. Exit
5. Reopen visualization

Let's visualize our data in Trackster

1. Create visualization



The screenshot shows the Galaxy web interface. The 'Visualization' menu is open, showing options for 'New Track Browser' and 'Saved Visualizations'. A yellow circle with the number '1' highlights the 'New Track Browser' option. The main data table displays genomic data for chromosome 1 (chr1) across various genes and coordinates. The table has columns for chromosome, coordinates, gene names, and counts. The right sidebar shows a history of datasets, including 'GCC Workshop', '12: Gene fusions', '11: Cell line mpileup', '10: UCSC Main on Human: refFlat (genome)', and '9: HPAC Transcriptome (KRAS, STK11, and ERBB)'. The user 'Jeremy' is logged in, and the interface is using 1.4 MB of memory.

1	2	3	4	5	6	7	8	9	10	11
chr1	11873	14409	DDX11L1	0	+					354,109,1189,
chr1	14361	29370	WASH7P	0	-	29370	29370	0	11	468,69,152,159,198,136,
chr1	17368	17436	MIR6859-1	0	-	17436	17436	0	1	68,
chr1	17368	17436	MIR6859-2	0	-	17436	17436	0	1	68,
chr1	34610	36081	FAM138A	0	-	36081	36081	0	3	564,205,361,
chr1	34610	36081	FAM138F	0	-	36081	36081	0	3	564,205,361,
chr1	69090	70008	OR4F5	0	+	69090	70008	0	1	918,
chr1	134772	140566	LOC729737	0	-	140566	140566	0	3	4924,58,492,
chr1	323891	328581	LOC100133331	0	+	328581	328581	0	4	169,58,2500,1546,
chr1	323891	328581	LOC100132287	0	+	328581	328581	0	3	169,58,4143,
chr1	323891	328581	LOC100132062	0	+	328581	328581	0	3	169,58,4143,
chr1	367658	368597	OR4F29	0	+	367658	368597	0	1	939,
chr1	367658	368597	OR4F3	0	+	367658	368597	0	1	939,

Let's visualize our data in Trackster

1. Create visualization

New Visualization

Browser name:
Workshop

Reference genome build (dbkey):
Mycobacterium sp. JLS (16079) ▲

hg19| [Custom Build](#)

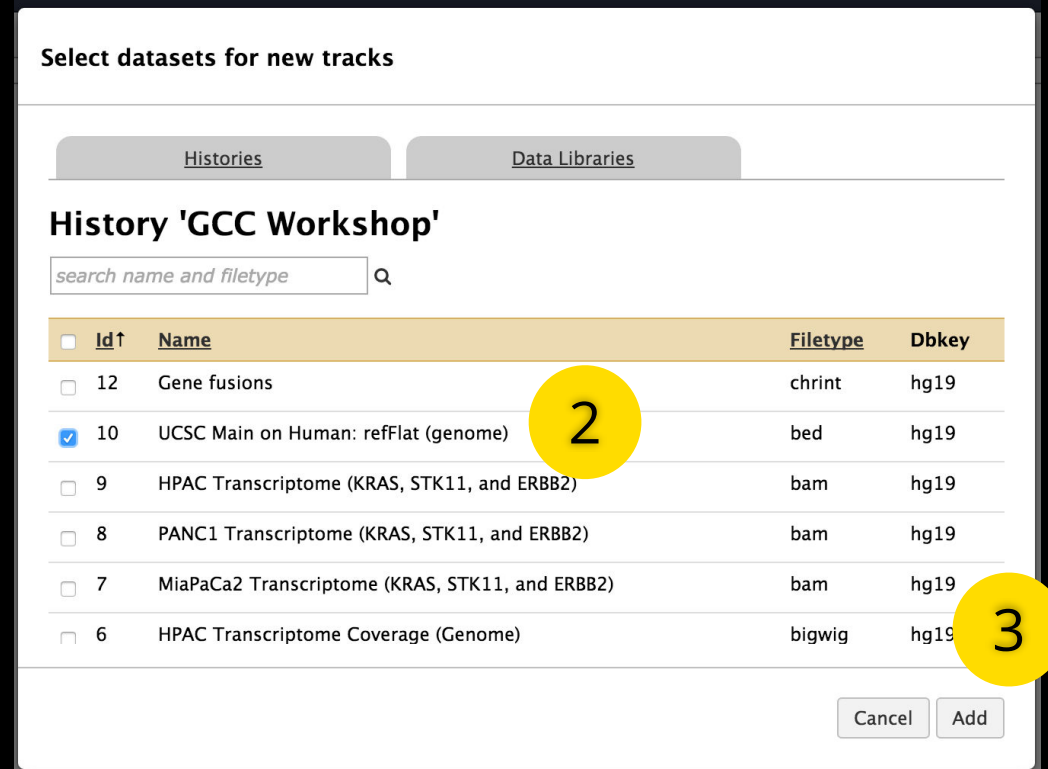
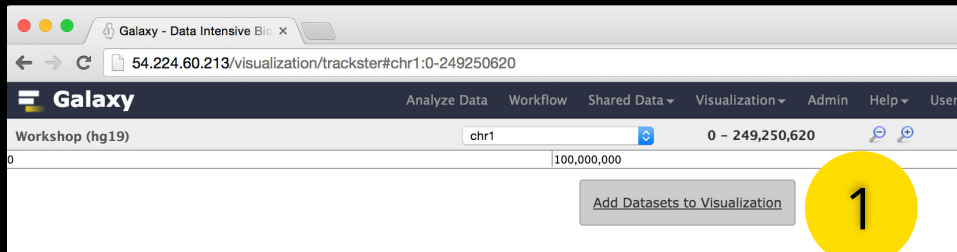
- (Homo_sapiens_nuHg19_mtrCRS)
- hg19Haplotypes Feb. 2009 (GRCh37/hg19Haps) (hg19Haps)
- Human Feb. 2009 (GRCh37/hg19) (hg19)**
- Hg19 lgg inversion Mar 2013 (1/hg19LggInv) (hg19LggInv)
- GRCh37.p2 Aug. 2009 (GRCh37.p2/hg19Patch2) (hg19Patch2)

Cancel Create

Let's visualize our data in Trackster

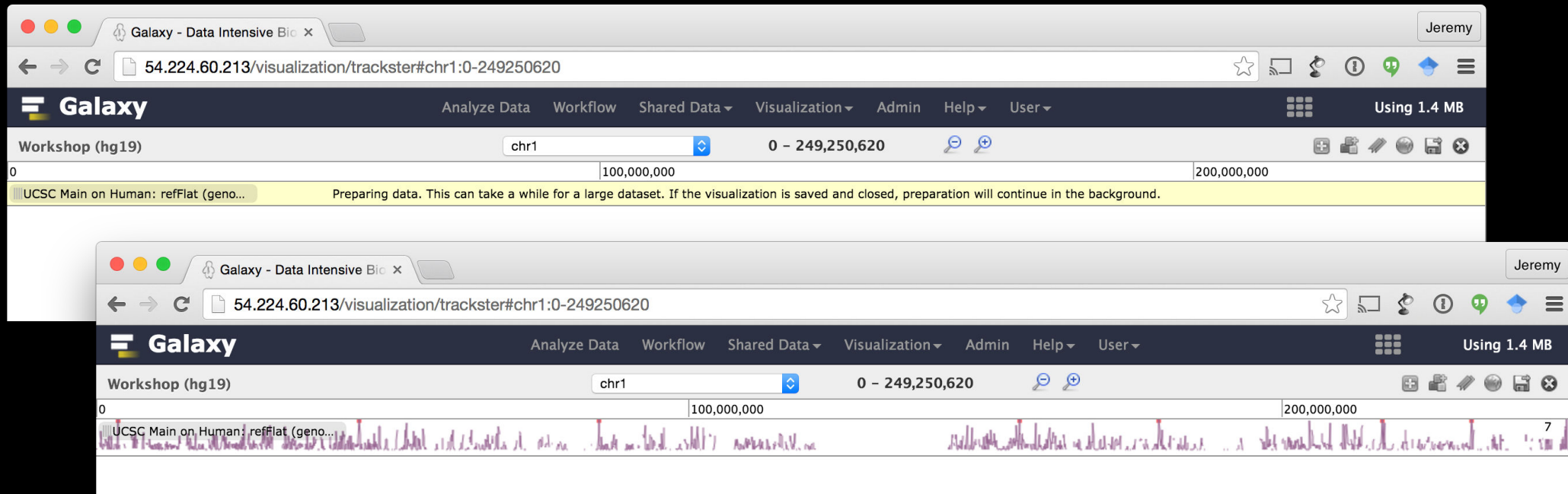
1. Create visualization

2. Add gene annotation (RefSeq)



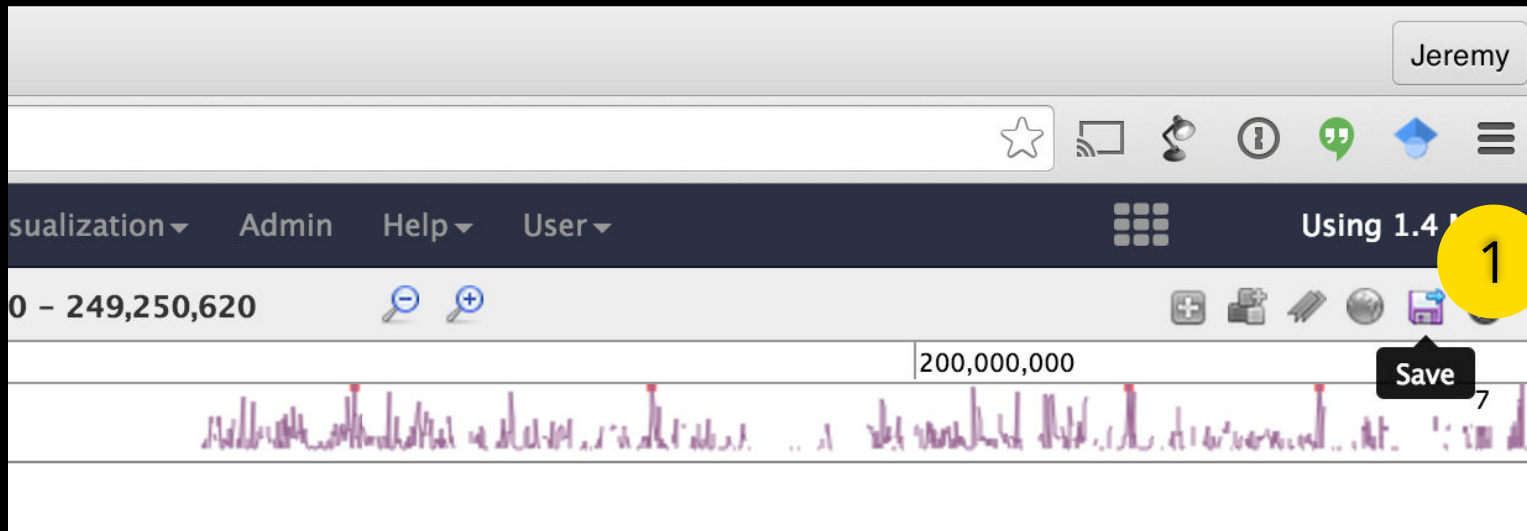
Let's visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)

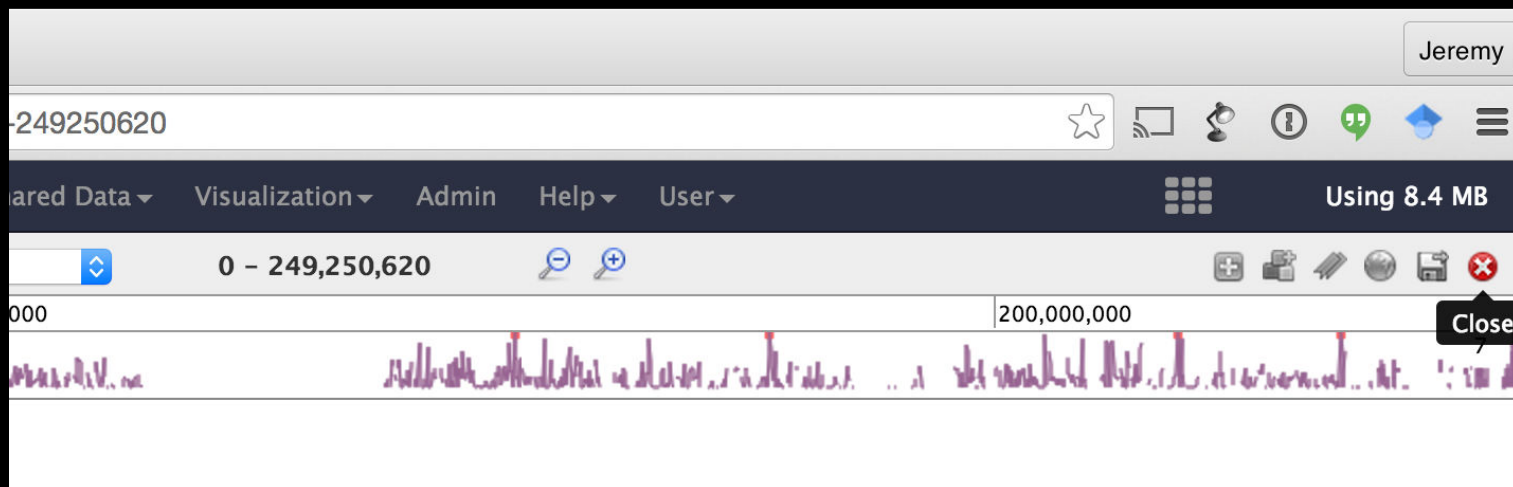


Let's visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)
3. Save visualization



Let's visualize our data in Trackster



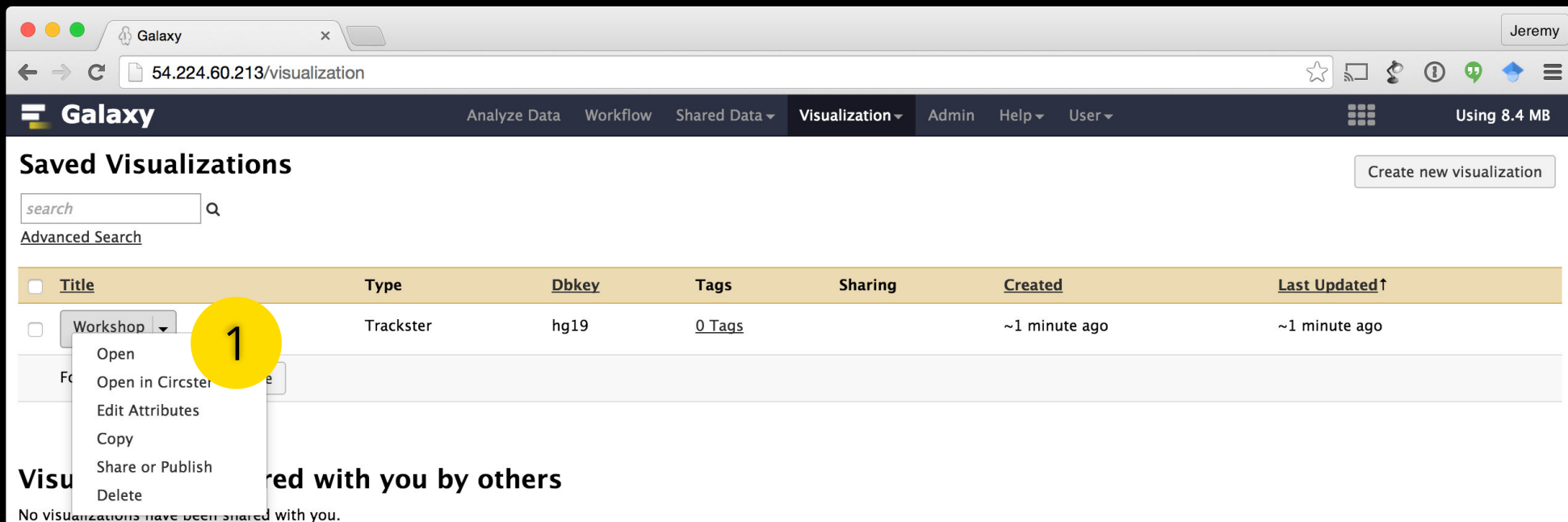
4. Exit

The screenshot shows the Galaxy web interface. The address bar displays the URL '54.224.60.213/visualization'. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User', along with a 'Using 8.4 MB' status. Below the navigation bar, the 'Saved Visualizations' section is displayed, featuring a search bar and a table of saved visualizations.

Title	Type	Dbkey	Tags	Sharing	Created	Last Updated ↑
Workshop	Trackster	hg19	0 Tags		~1 minute ago	~1 minute ago

For 0 selected items: [Delete](#)

Let's visualize our data in Trackster



The screenshot shows the Galaxy web interface. The browser address bar displays "54.224.60.213/visualization". The Galaxy navigation bar includes links for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". The "Visualization" section is active, showing "Saved Visualizations". A search bar with the placeholder "search" and a "Create new visualization" button are present. Below the search bar is a table of saved visualizations. The table has columns: Title, Type, Dbkey, Tags, Sharing, Created, and Last Updated. A yellow circle with the number 1 highlights the "Workshop" dropdown menu, which is open, showing options: Open, Open in Circster, Edit Attributes, Copy, Share or Publish, and Delete. Below the table, there is a section titled "Visualizations shared with you by others" and a message: "No visualizations have been shared with you."

Title	Type	Dbkey	Tags	Sharing	Created	Last Updated
Workshop	Trackster	hg19	0 Tags		~1 minute ago	~1 minute ago

5. Reopen visualization

Behind the Scenes

Galaxy is indexing datasets for

- ✦ viewing large genomic regions (coverage plots)
- ✦ viewing small genomic regions (getting individual data points)
- ✦ feature names and locations

Indexes is the primary way that big datasets are visualized quickly

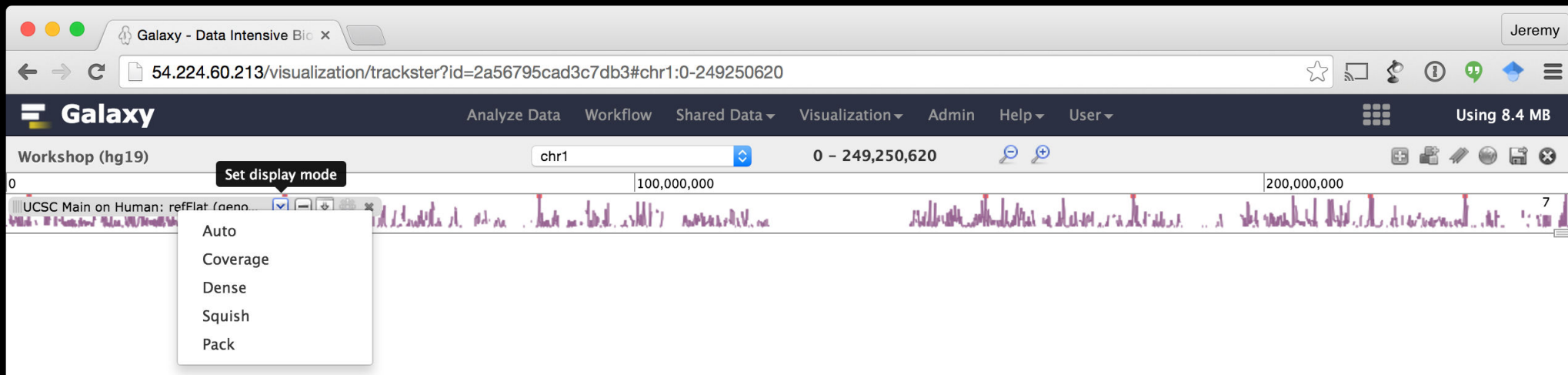
Display Modes

Tracks can be displayed differently

- ✦ coverage to individual features
- ✦ similar language to UCSC

Let's try different modes

- ✦ this is fast because data is sent from Galaxy server and rendered in your Web browser



The screenshot shows the Galaxy web interface. The browser address bar displays the URL: `54.224.60.213/visualization/trackster?id=2a56795cad3c7db3#chr1:0-249250620`. The Galaxy logo is in the top left, and the user name "Jeremy" is in the top right. The main navigation bar includes links for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", and "User". The current view is a "Workshop (hg19)" showing a track for "chr1" with a range of "0 - 249,250,620". A dropdown menu labeled "Set display mode" is open, showing the following options: "Auto", "Coverage", "Dense", "Squish", and "Pack". The track itself displays a genomic visualization with red and blue bars representing data coverage.

Searching

Can search for named features such as gene annotations

- ✦ BED, GFF/GTF

Let's try searching for a gene: ERBB2

The screenshot shows the Galaxy web interface. The browser address bar displays the URL: `54.224.60.213/visualization/trackster?id=2a56795cad3c7db3#chr1:0-249250620`. The Galaxy header includes navigation links: Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. The main content area shows a track visualization for chromosome 1. A search bar at the top of the track area contains the text "erbb". Below the search bar, a dropdown menu lists the search results: ERBB2, ERBB2IP, ERBB3, and ERBB4. The track visualization itself shows a genomic track with various features and annotations, including gene names like ERBB2, ERBB2IP, ERBB3, and ERBB4. The track is labeled "Workshop (hg19)" and "erbb". The track visualization shows a genomic track with various features and annotations, including gene names like ERBB2, ERBB2IP, ERBB3, and ERBB4. The track is labeled "Workshop (hg19)" and "erbb".

Let's Call Variants

VarScan

- ✦ Sample names:
MiaPaCa2, PANC1, HPAC
- ✦ Run

Rename output: "Cell line variants"

Varscan for variant detection (Galaxy Tool Version 0.1) Options

Pileup dataset
11: Cell line mpileup **1**

Analysis type
single nucleotide variation

Minimum read depth
8

Minimum depth at a position to make a call

Minimum supporting reads
2

Minimum supporting reads at a position to make a call

Minimum base quality at a position to count a read
15

Minimum variant allele frequency threshold
0.01

Minimum frequency to call homozygote
0.75

p-value threshold for calling variants
0.99

Ignore variants with >90% support on one strand
no

sample_names
MiaPaCa2, PANC1, HPAC **2**

Separate sample names by comma. Leave blank to use default sample names.

3

Let's Assemble Transcripts

Cufflinks

- ✦ select transcriptome datasets
- ✦ run

Rename assembled transcripts for MiaPaCa2: "MiaPaCa2 Assembled Transcripts"

Cufflinks transcript assembly and FPKM (RPKM) estimates for RNA-Seq data (Galaxy Tool Version 2.2.1.0) Options

SAM or BAM file of aligned RNA-Seq reads

1 2

9: HPAC Transcriptome (KRAS, STK11, and ERBB2)
8: PANC1 Transcriptome (KRAS, STK11, and ERBB2)
7: MiaPaCa2 Transcriptome (KRAS, STK11, and ERBB2)
3: HPAC Exome (KRAS, STK11, and ERBB2)
2: PANC1 Exome (KRAS, STK11, and ERBB2)

This is a batch mode input field. A separate job will be triggered for each dataset.

Max Intron Length
300000
ignore alignments with gaps longer than this

Min Isoform Fraction
0.1
suppress transcripts below this abundance level

Pre mRNA Fraction
0.15
suppress intra-intronic transcripts below this level

Use Reference Annotation
No

Perform Bias Correction
No
Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

Use multi-read correct
No
Tells Cufflinks to do an initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome.

Apply length correction
Cufflinks Effective Length Correction
Mode of length normalization to transcript FPKM.

Set advanced Cufflinks options

Let's add data to Trackster

Add exome data for all cell lines and called variants...

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Workshop (hg19)

UCSC Main on Human: refFlat

1

Select datasets for new tracks

☒	14	Cell line variants	vcf	hg19
☐	13	Gene fusions	chrnt	hg19
☐	10	UCSC Main on Human: refFlat (genome)	bed	hg19
☐	9	HPAC Transcriptome (KRAS, STK11, and ERBB2)	bam	hg19
☐	8	PANC1 Transcriptome (KRAS, STK11, and ERBB2)	bam	hg19
☐	7	MiaPaCa2 Transcriptome (KRAS, STK11, and ERBB2)	bam	hg19
☐	6	HPAC Transcriptome Coverage (Genome)	bigwig	hg19
☐	5	PANC1 Transcriptome Coverage (Genome)	bigwig	hg19
☐	4	MiaPaCa2 Transcriptome Coverage (Genome)	bigwig	hg19
☒	3	HPAC Exome (KRAS, STK11, and ERBB2)	bam	hg19
☒	2	PANC1 Exome (KRAS, STK11, and ERBB2)	bam	hg19
☒	1	MiaPaCa2 Exome (KRAS, STK11, ERBB2)	bam	hg19

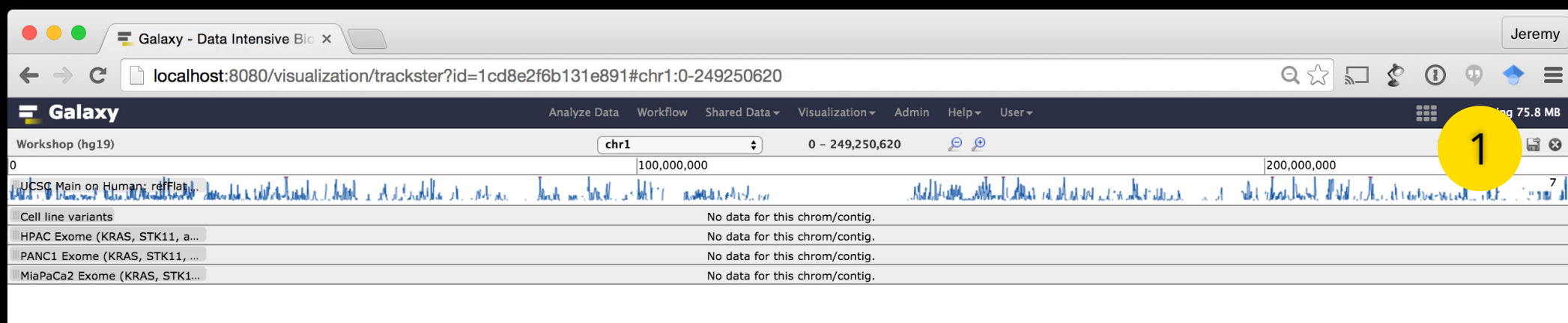
2

3

Cancel Add

Let's add data to Trackster

Add exome data for all cell lines and called variants...
...but where is our data?

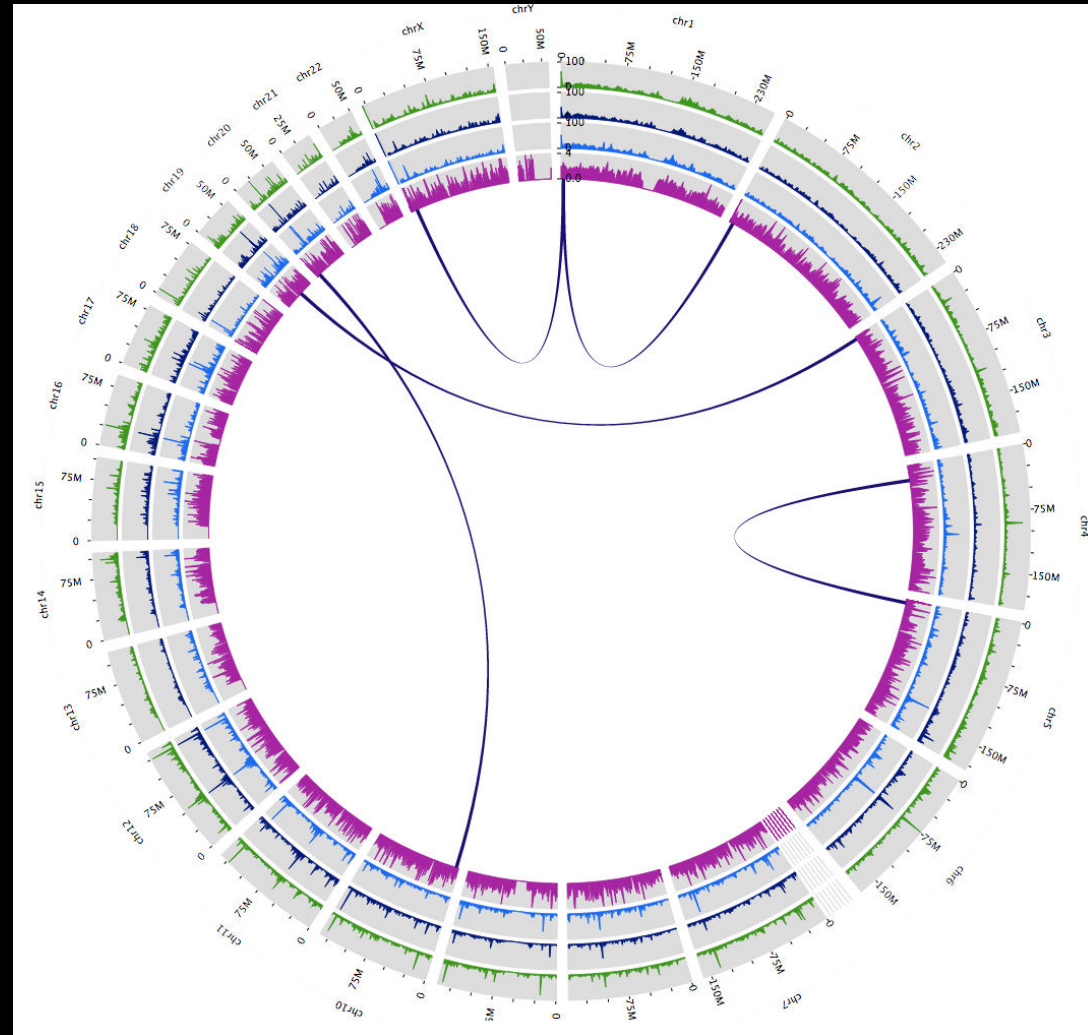


Save again

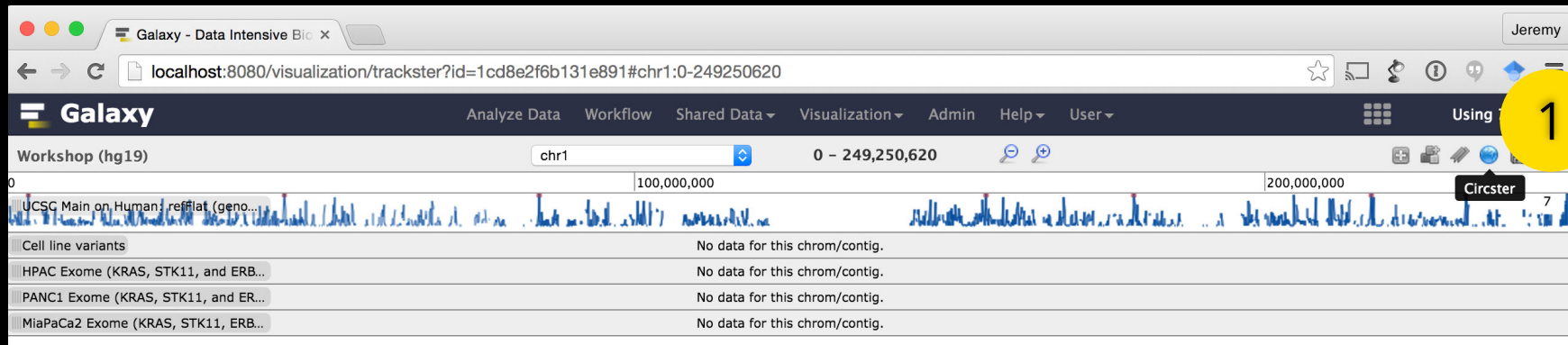
Circster

Interactive Circos
plot

Whole genome
view with structural
variation



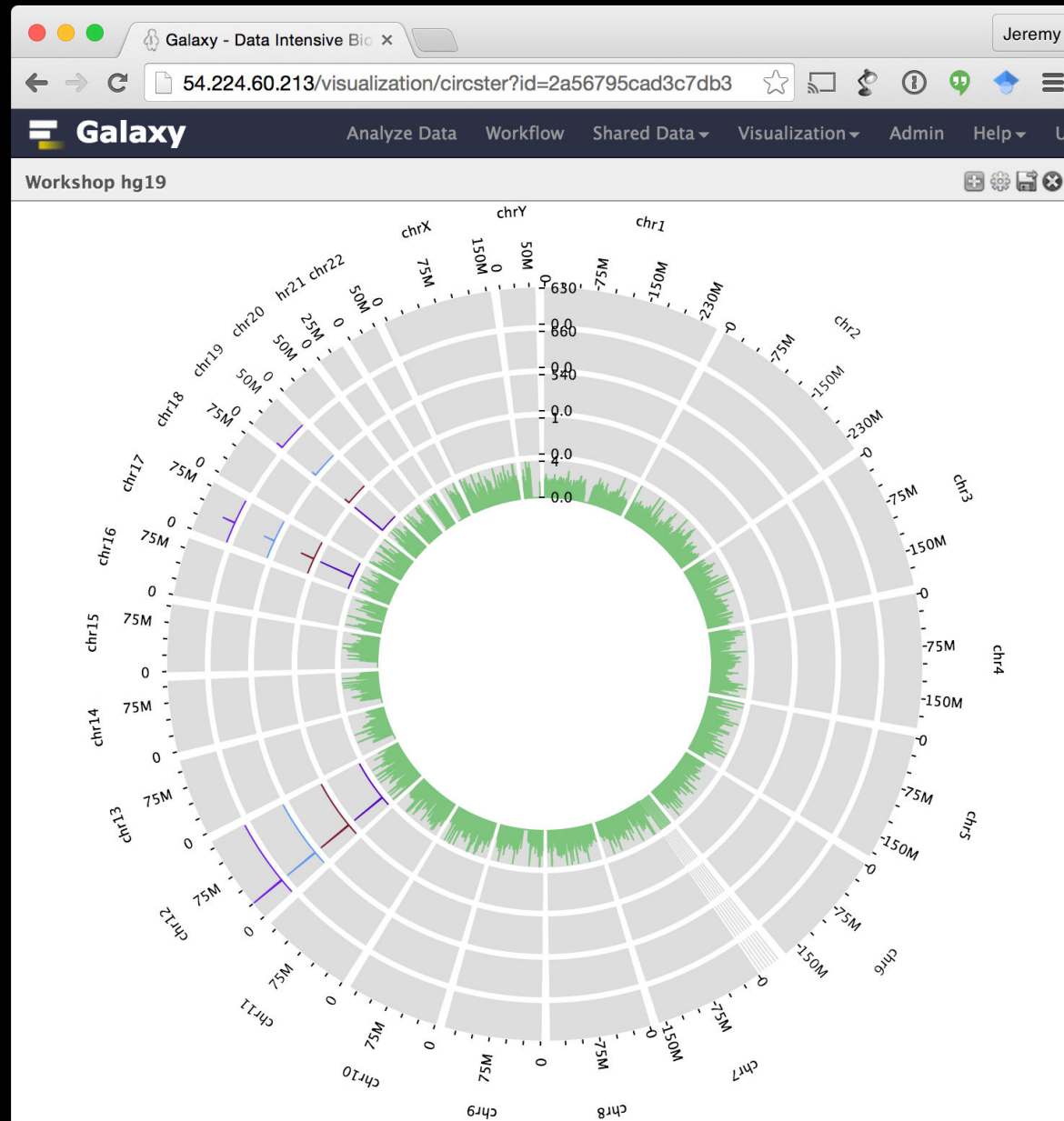
Let's view our data in Circster



The screenshot shows the 'Saved Visualizations' page in the Galaxy web interface. The page lists three saved visualizations, all of which are 'Trackster' type and associated with 'hg19'. The first visualization is highlighted with a yellow circle and the number '1'. The context menu for this visualization is open, showing options: 'Open', 'Open in Circster', 'Edit Attributes', 'Copy', 'Share or Publish', and 'Delete'. The 'Open in Circster' option is selected.

Title	Type	Dbkey	Tags	Sharing	Created	Last Updated†
Workshop	Trackster	hg19	0 Tags		~16 minutes ago	~2 minutes ago
Open	Trackster	hg19	0 Tags		Jun 18, 2015	Jun 18, 2015
Open in Circster	Trackster	hg19	0 Tags		Mar 31, 2015	Mar 31, 2015
Edit Attributes						
Copy						
Share or Publish						
Delete						

Let's view our data in Circster

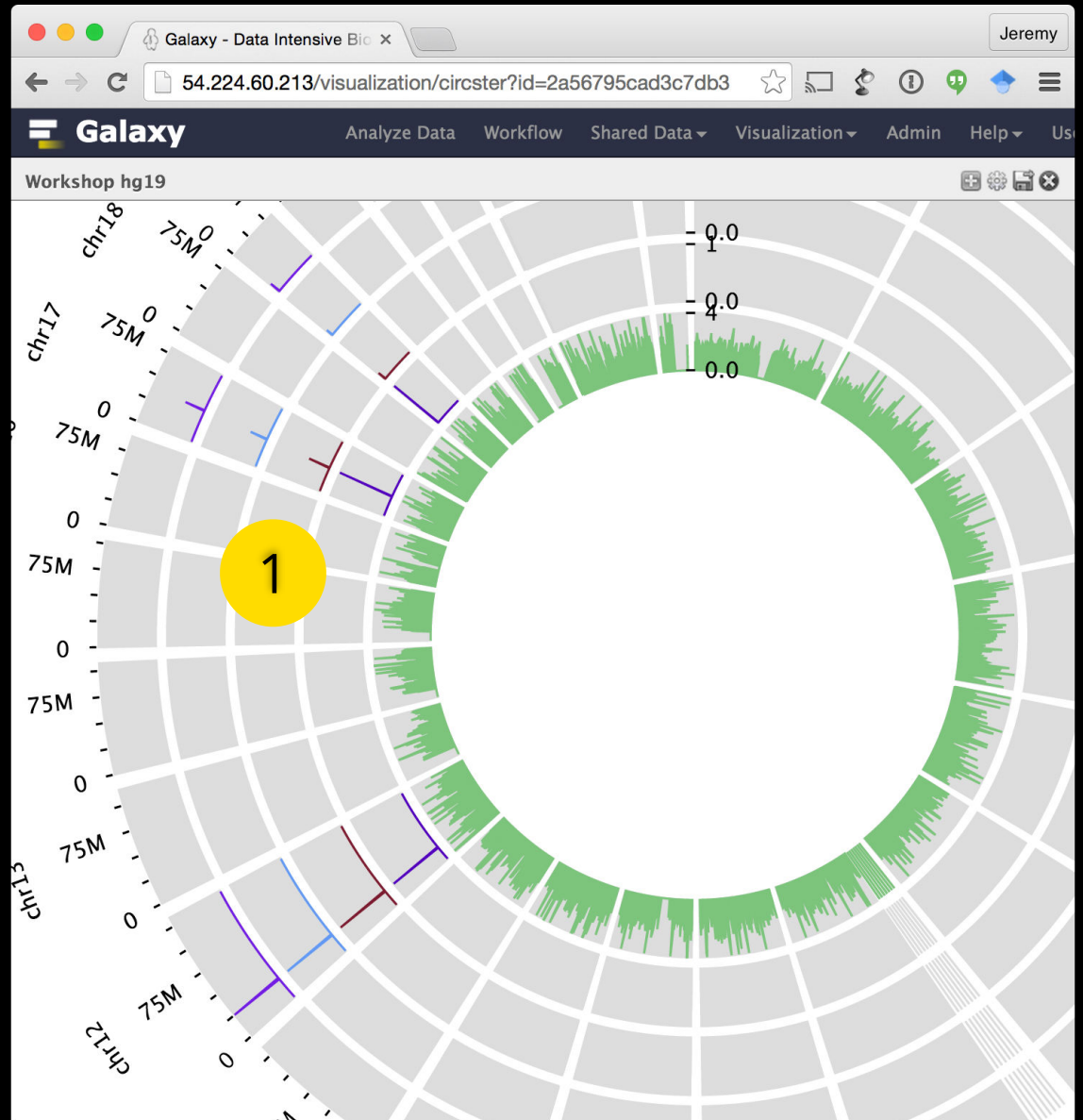


Let's view our data in Circster

Double-click or use trackpad to zoom in

Drag around using mouse/trackpad

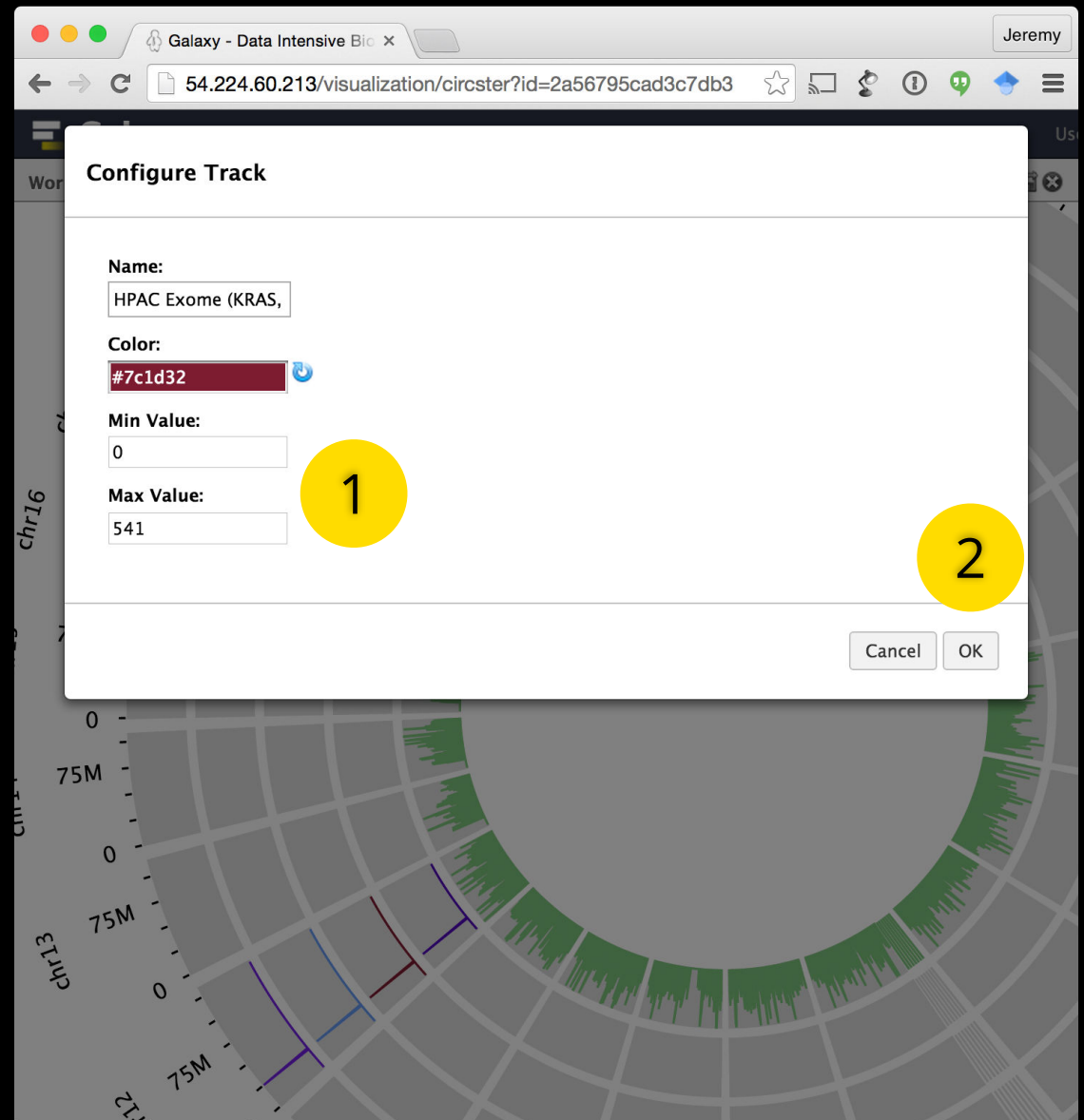
What do we see?



Let's view our data in Circster

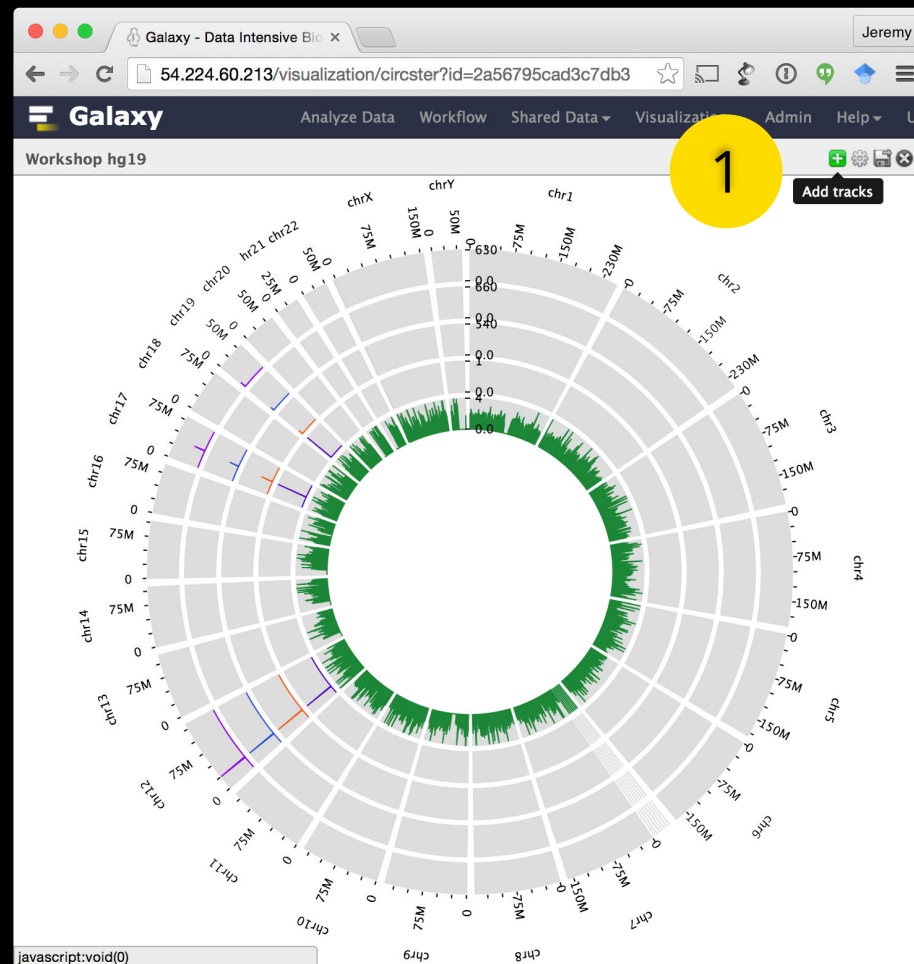
Change min/
max by clicking
on labels

What do we see?



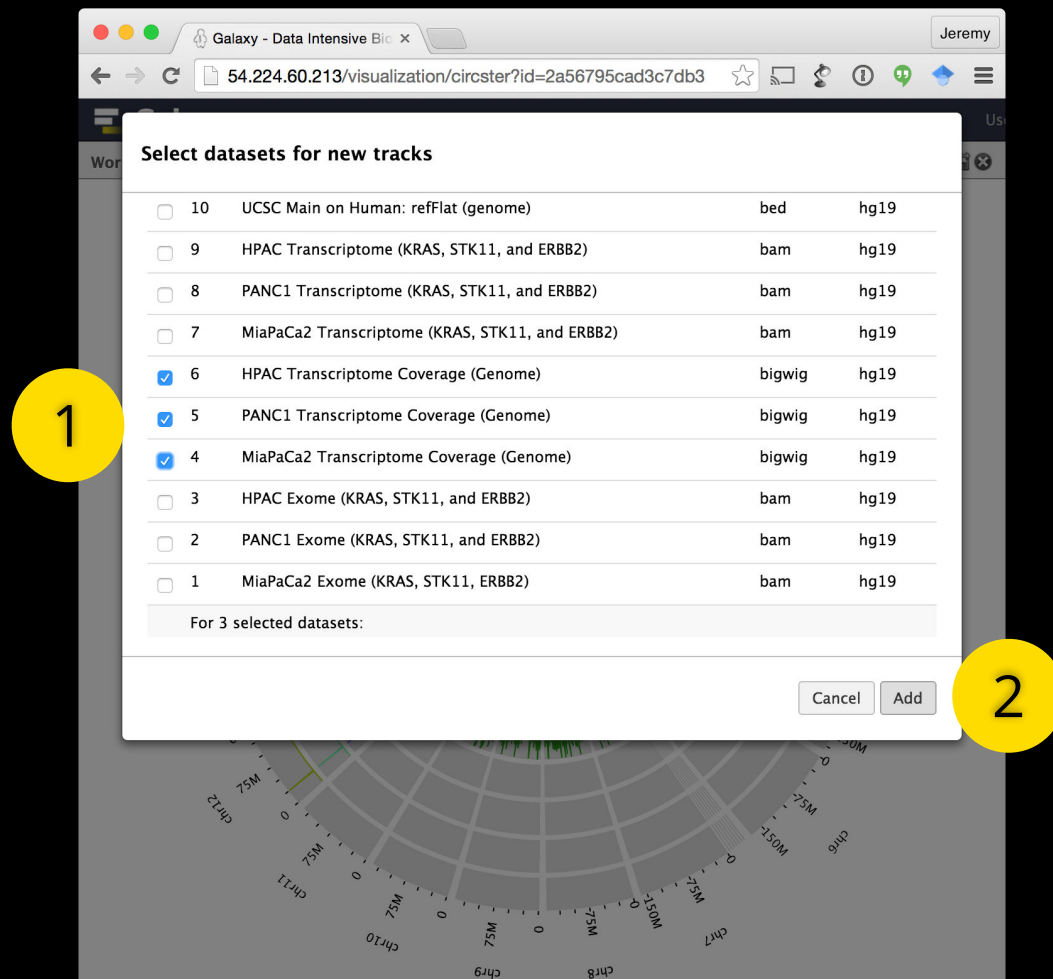
Let's add data to Circster and adjust options

1. Add transcriptome coverage data



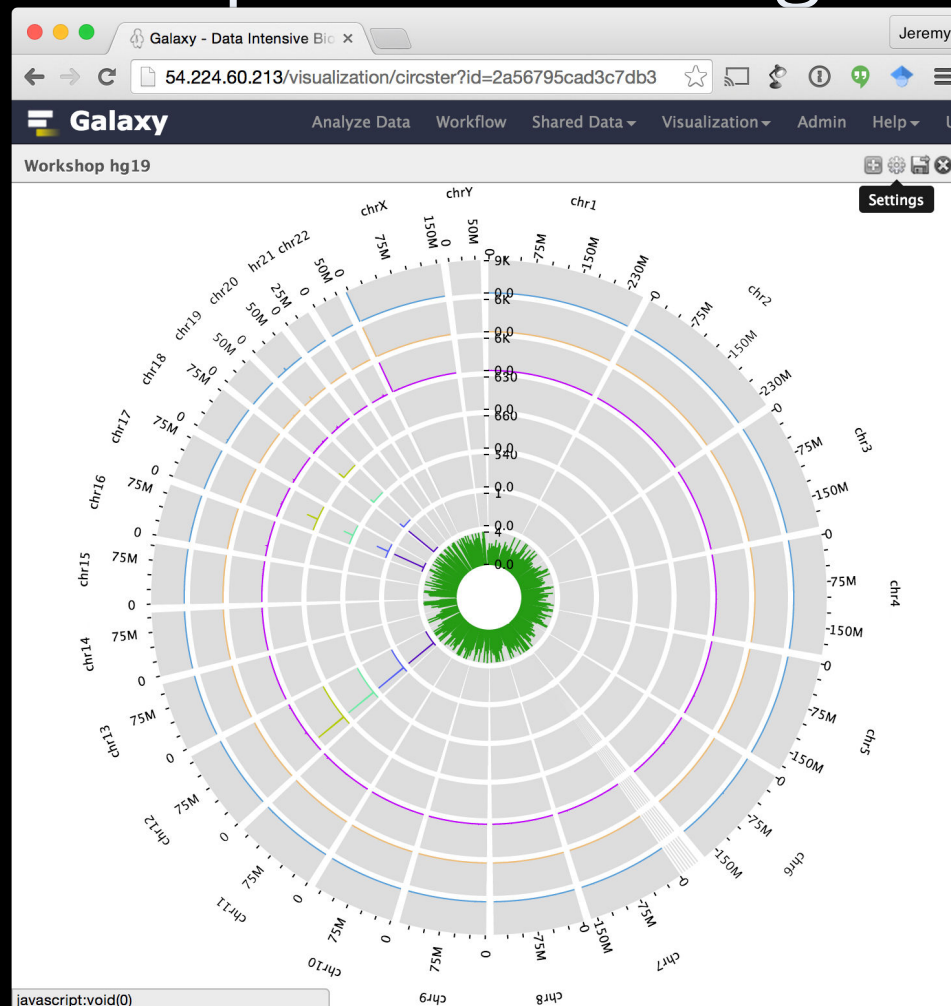
Let's add data to Circster and adjust options

1. Add transcriptome coverage data



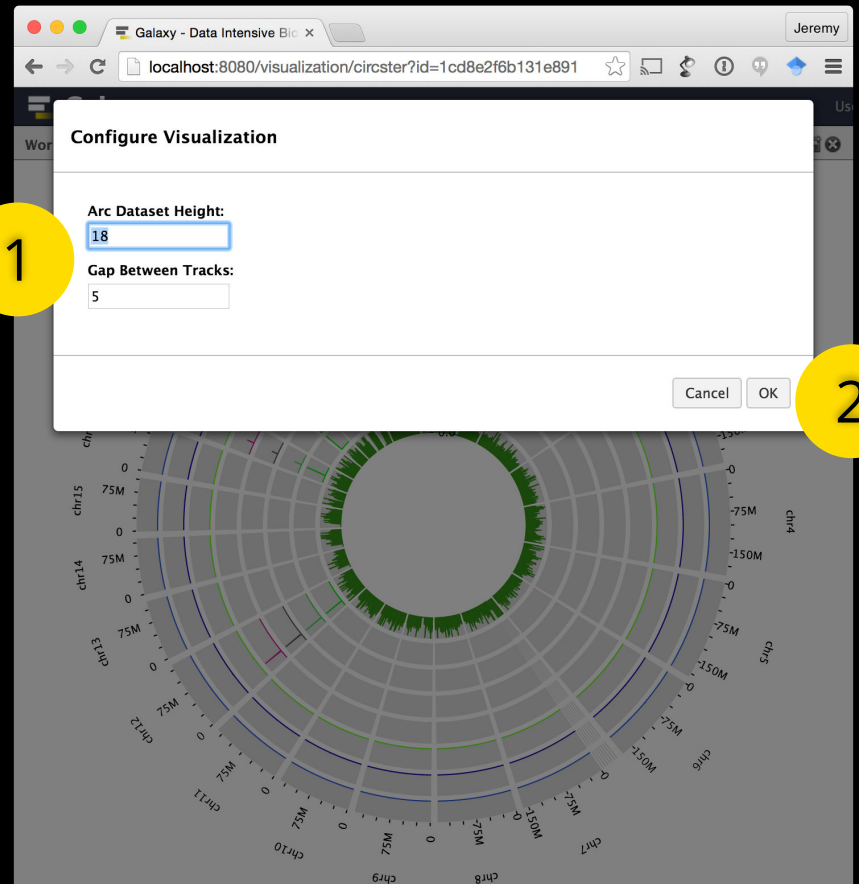
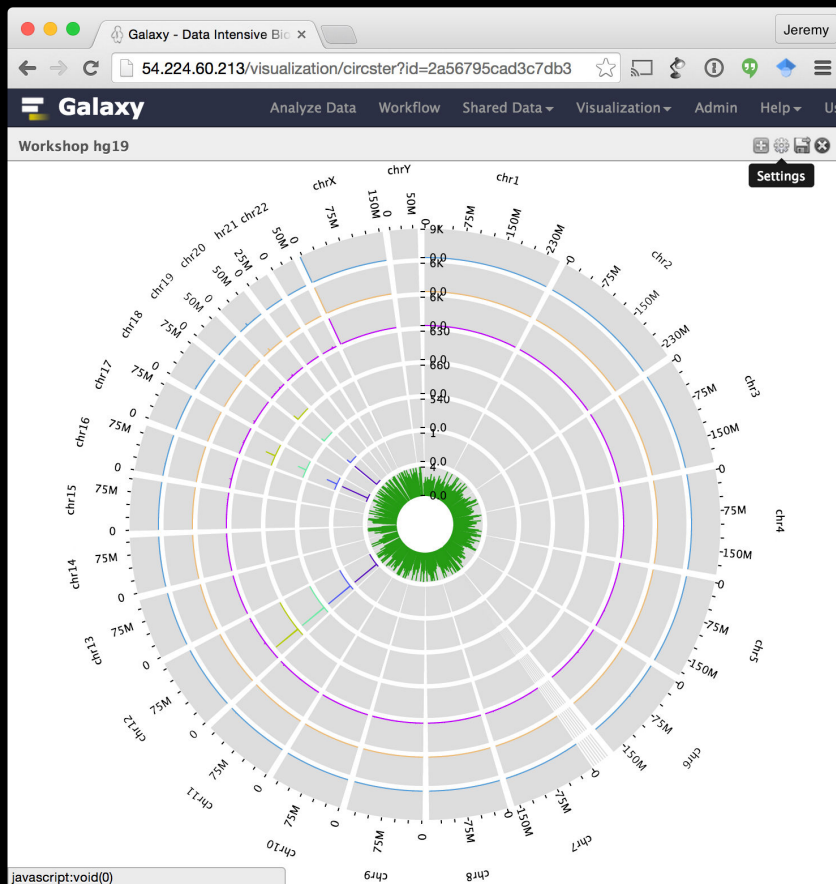
Let's add data to Circster and adjust options

1. Add transcriptome coverage data



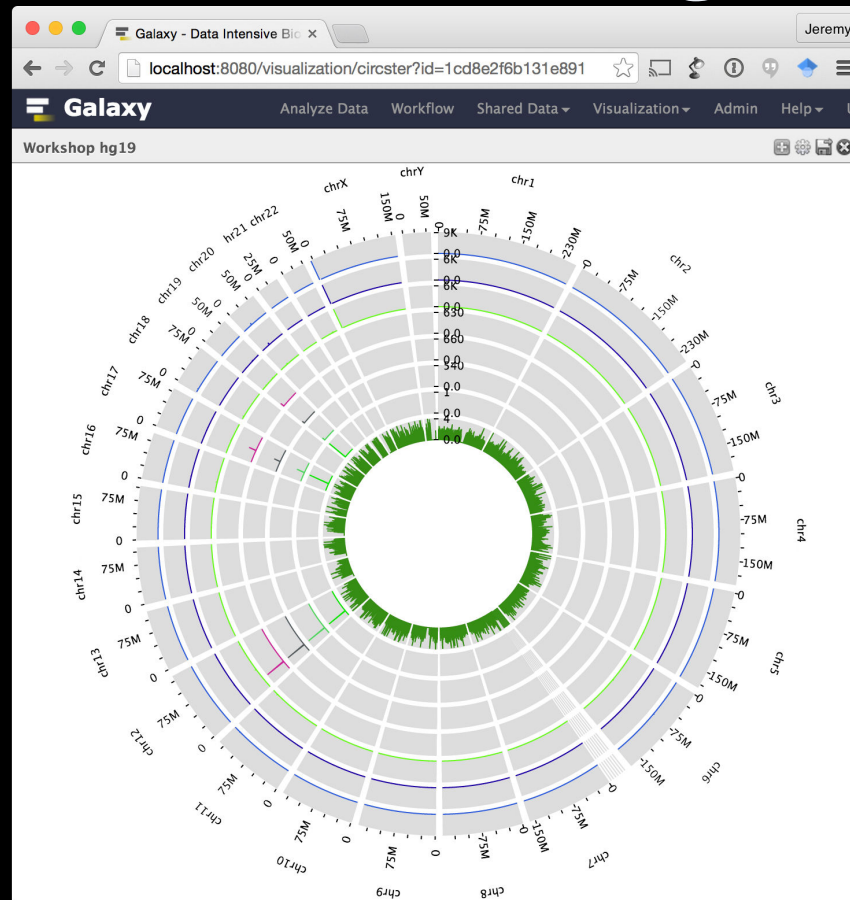
Let's add data to Circster and adjust options

2. Change arc dataset height



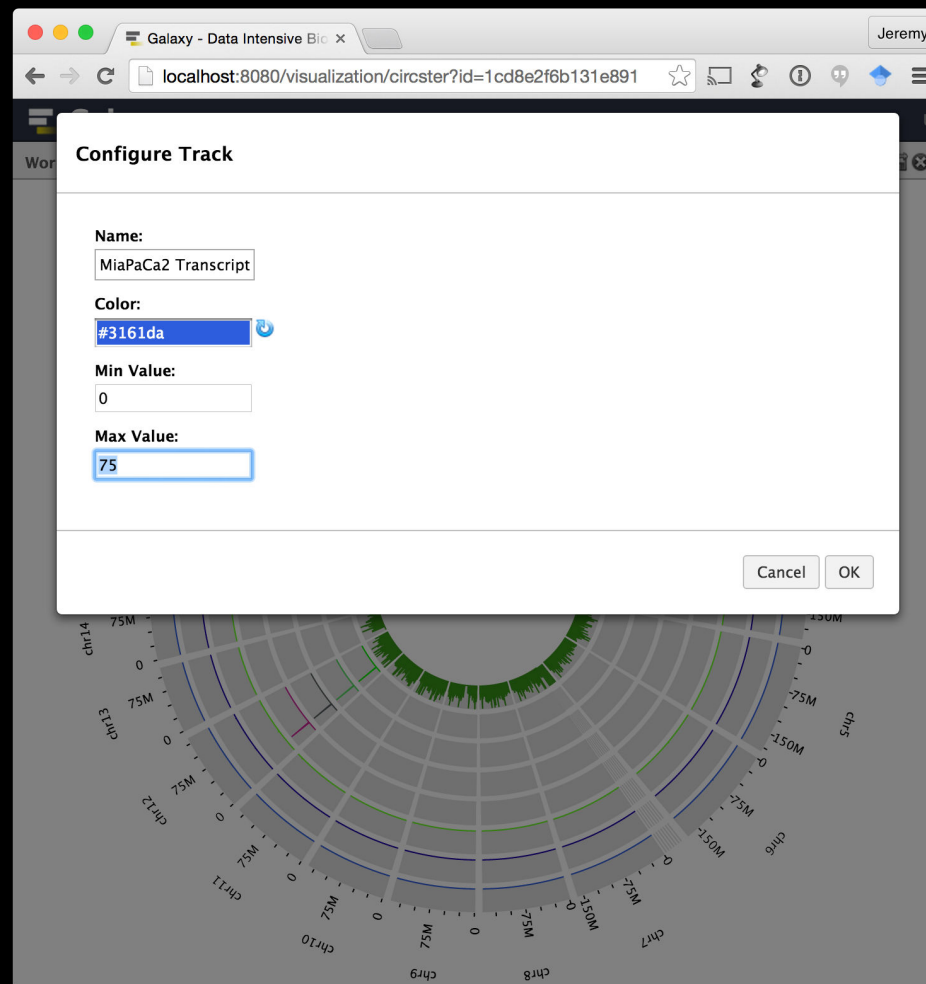
Let's add data to Circster and adjust options

2. Change arc dataset height



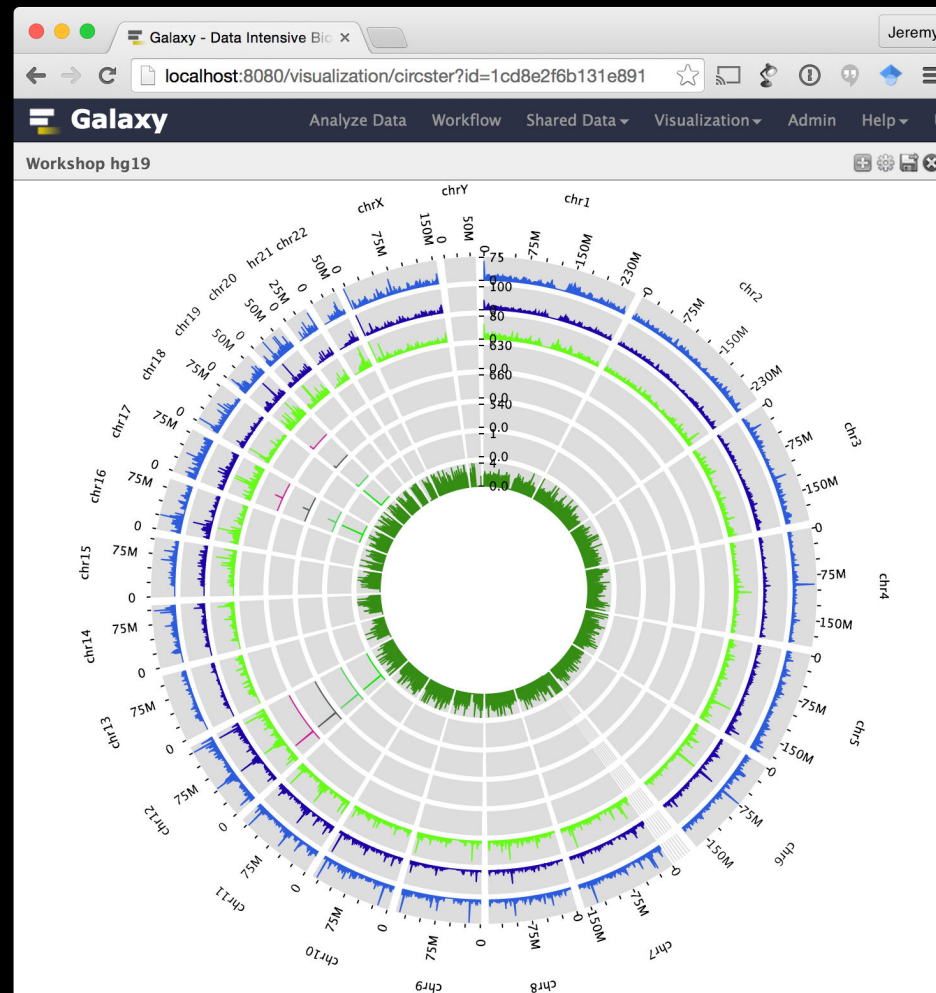
Let's add data to Circster and adjust options

3. Change max for tracks



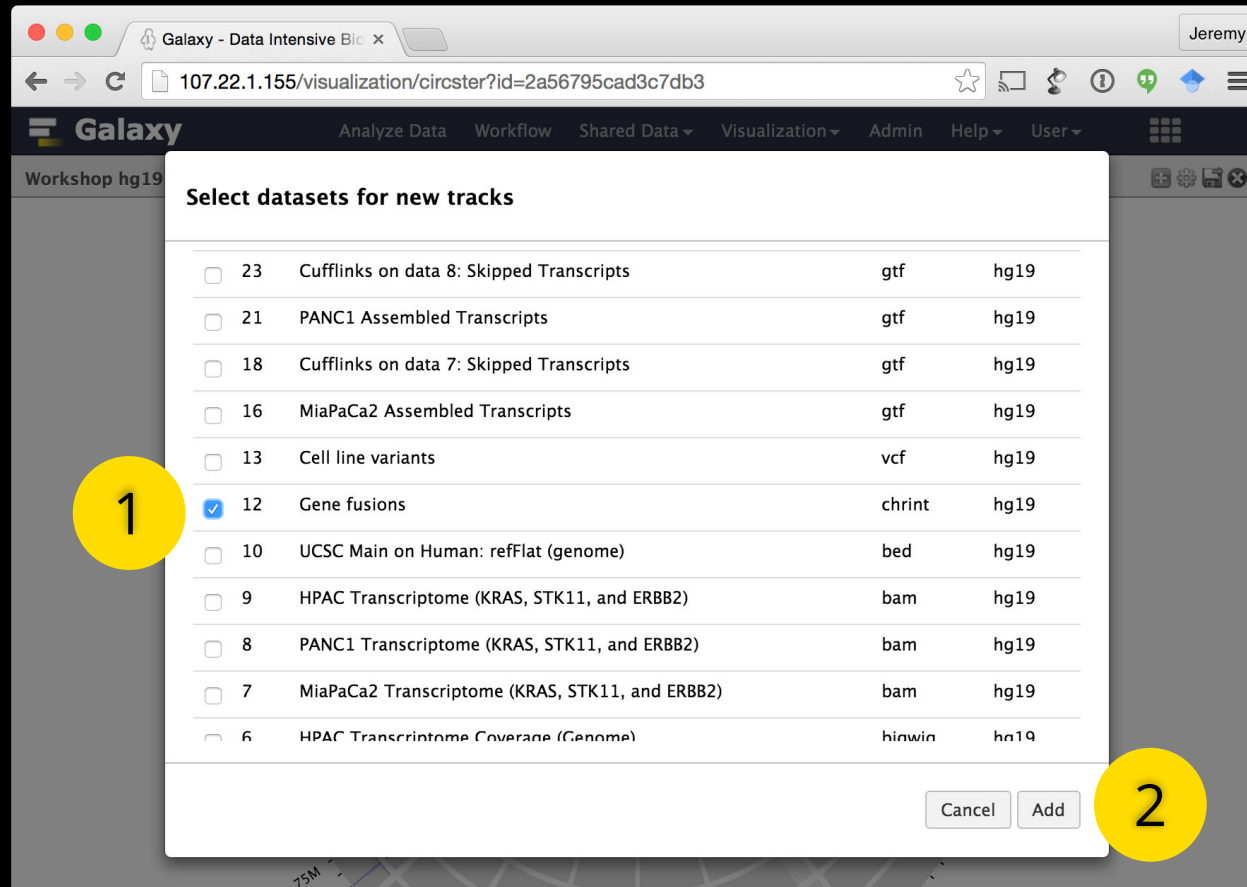
Let's add data to Circster and adjust options

3. Change max for tracks *what do we see?*



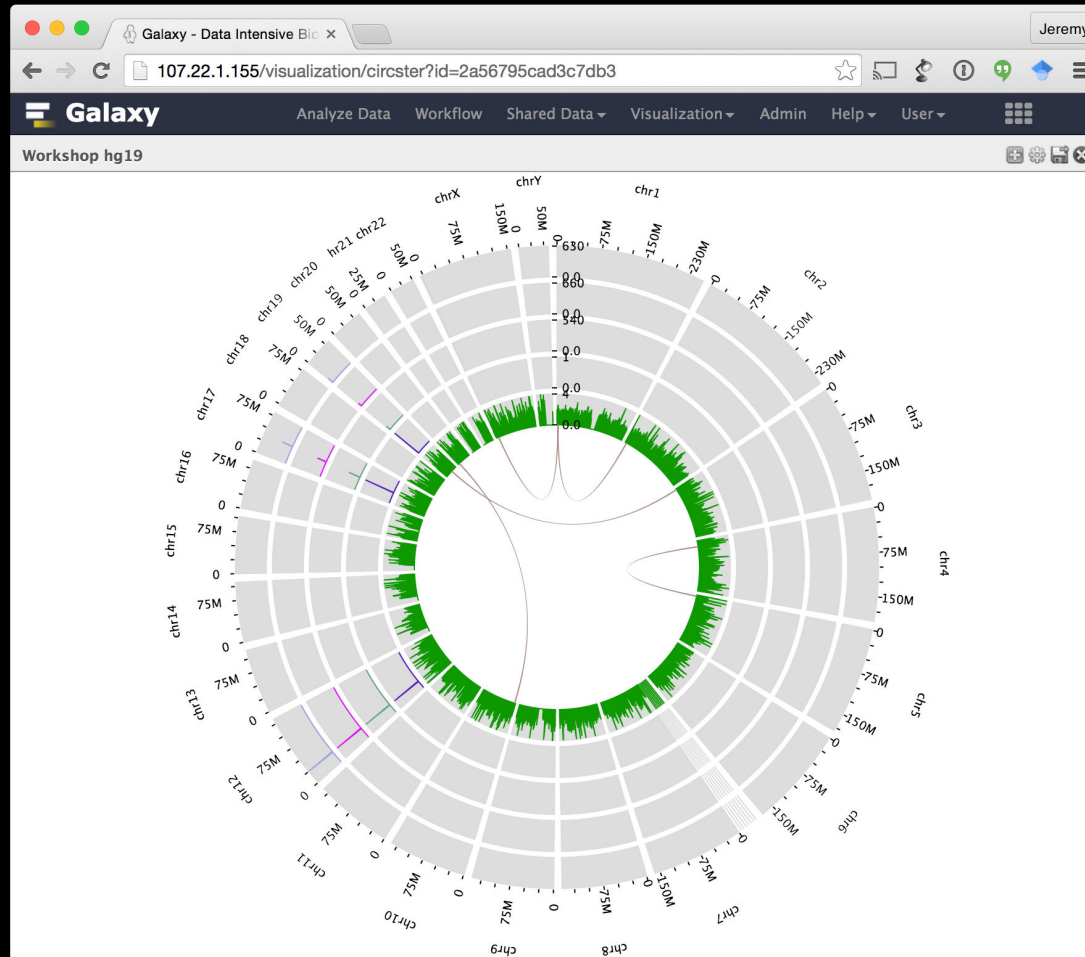
Let's add data to Circster and adjust options

Add gene fusions



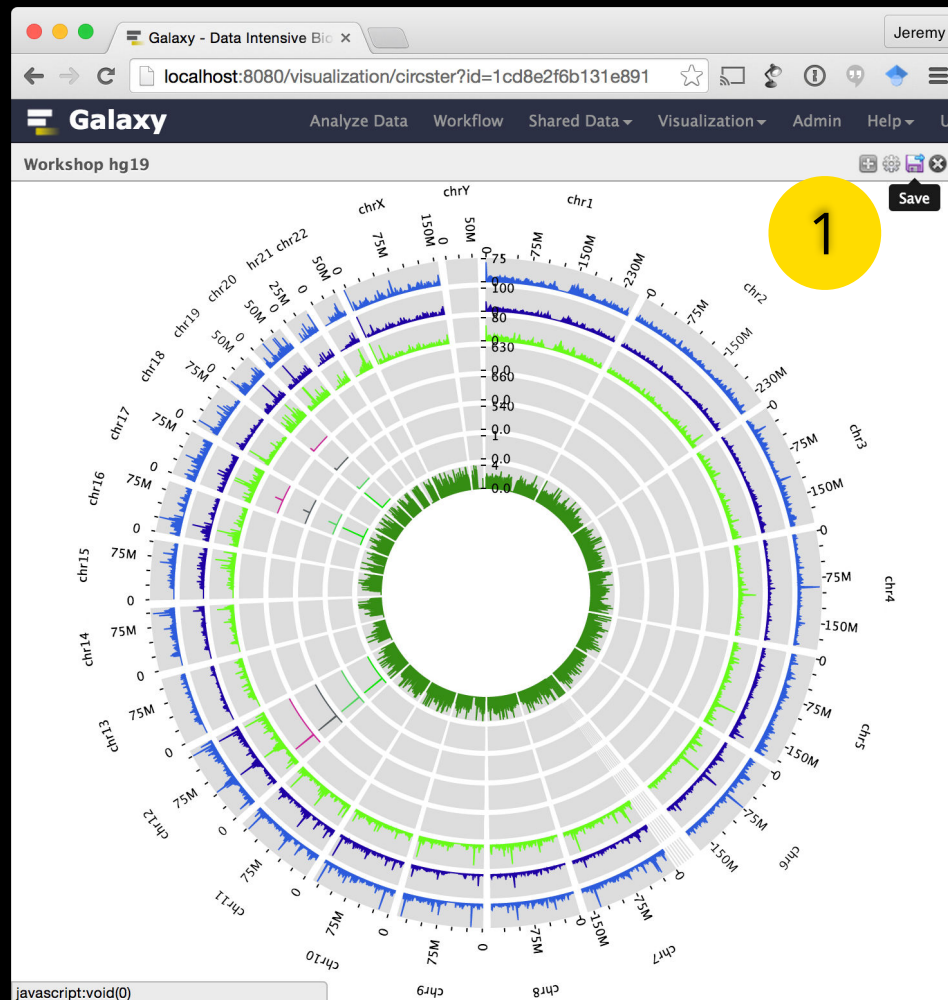
Let's add data to Circster and adjust options

Add gene fusions



Let's add data to Circster and adjust options

4. Save visualization

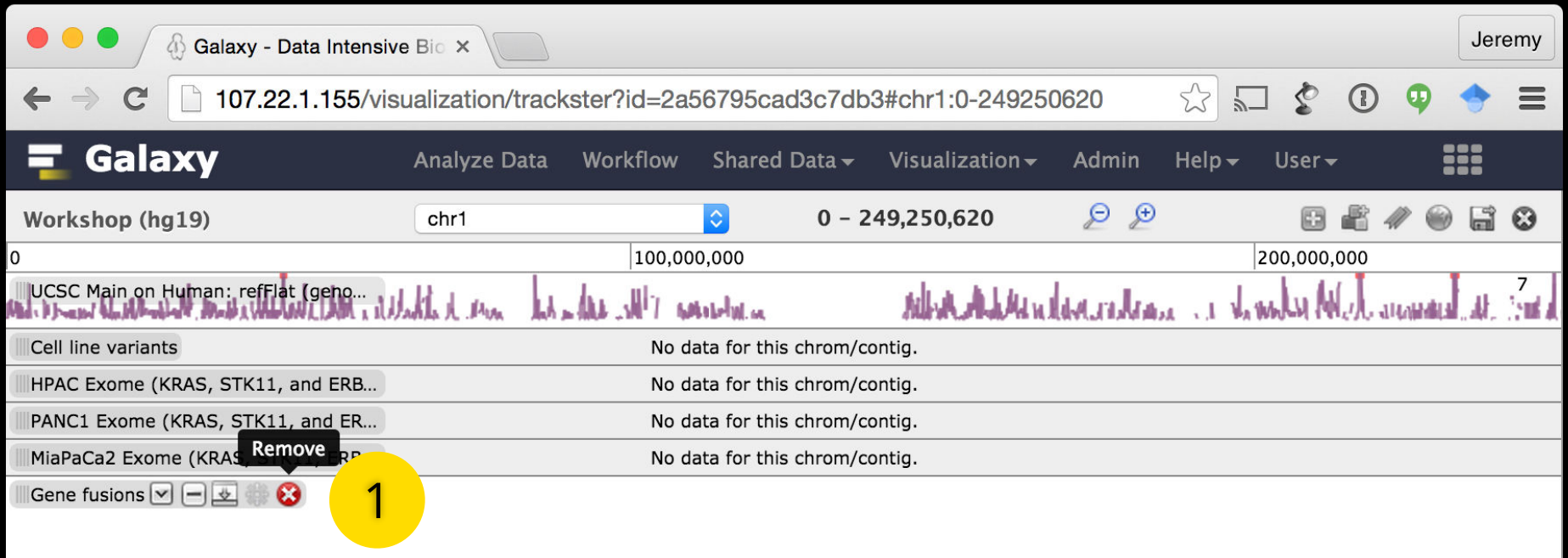


Back to Trackster: Rainbow Track for Coverage

1. Remove gene fusions track
2. Navigate to ERBB2 gene
3. Create group
4. Add transcriptome coverage tracks to group
5. Create composite track
6. Adjust max
7. *what do we see?*

Back to Trackster: Rainbow Track for Coverage

1. Remove gene fusions track



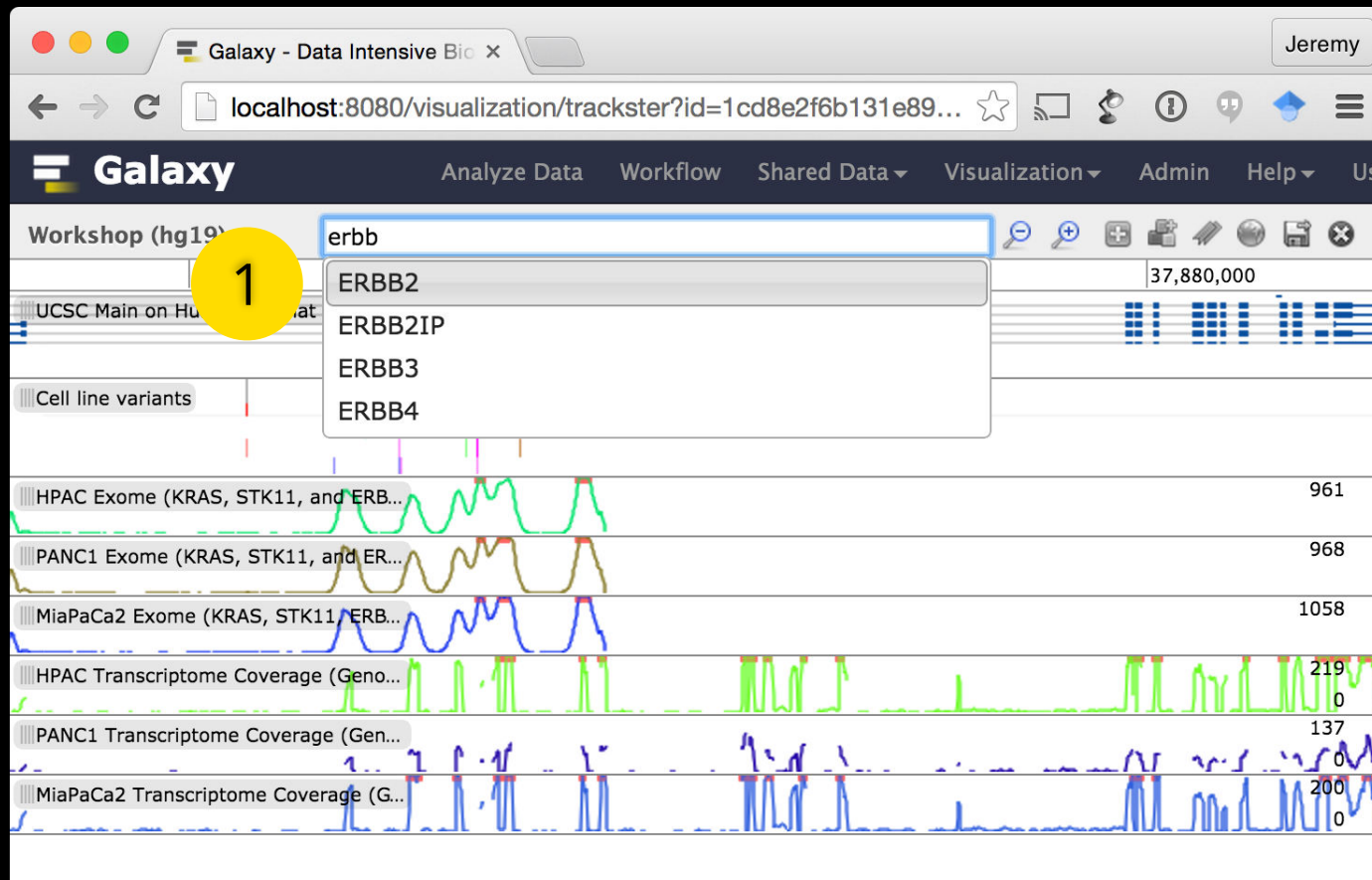
The screenshot shows the Galaxy web interface with the Trackster visualization. The browser address bar displays the URL: `107.22.1.155/visualization/trackster?id=2a56795cad3c7db3#chr1:0-249250620`. The interface includes a navigation bar with links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. The main content area shows a track visualization for chromosome 1 (chr1) with a zoom range of 0 to 249,250,620. The tracks listed are:

- UCSC Main on Human: refFlat (geho...)
- Cell line variants
- HPAC Exome (KRAS, STK11, and ERB...)
- PANC1 Exome (KRAS, STK11, and ER...)
- MiaPaCa2 Exome (KRAS, STK11, and ER...)
- Gene fusions

The 'Gene fusions' track is highlighted with a red 'X' icon and a yellow circle with the number 1, indicating it should be removed. A 'Remove' button is also visible next to the track name.

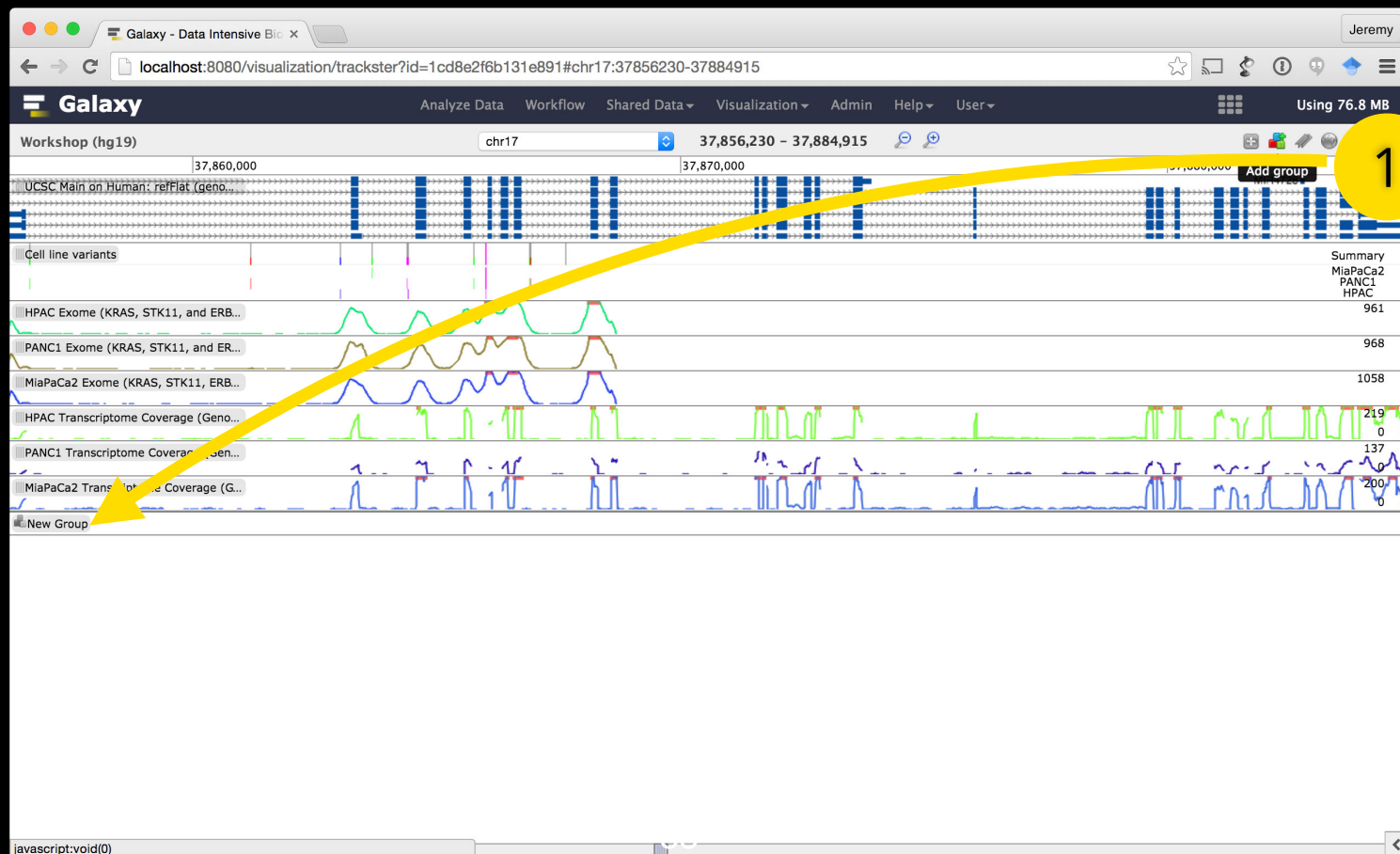
Back to Trackster: Rainbow Track for Coverage

2. Navigate to ERBB2 gene



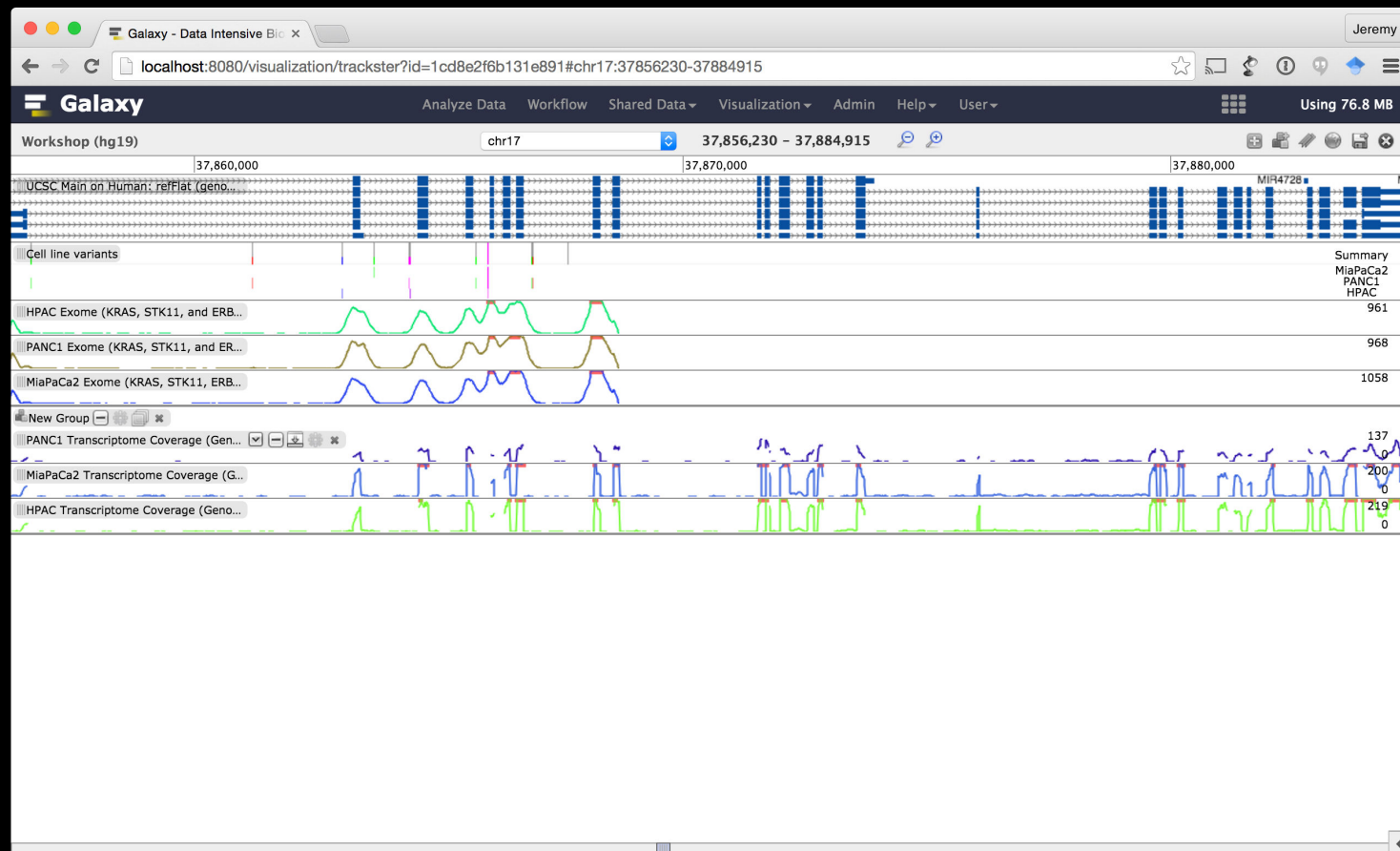
Back to Trackster: Rainbow Track for Coverage

3. Create group



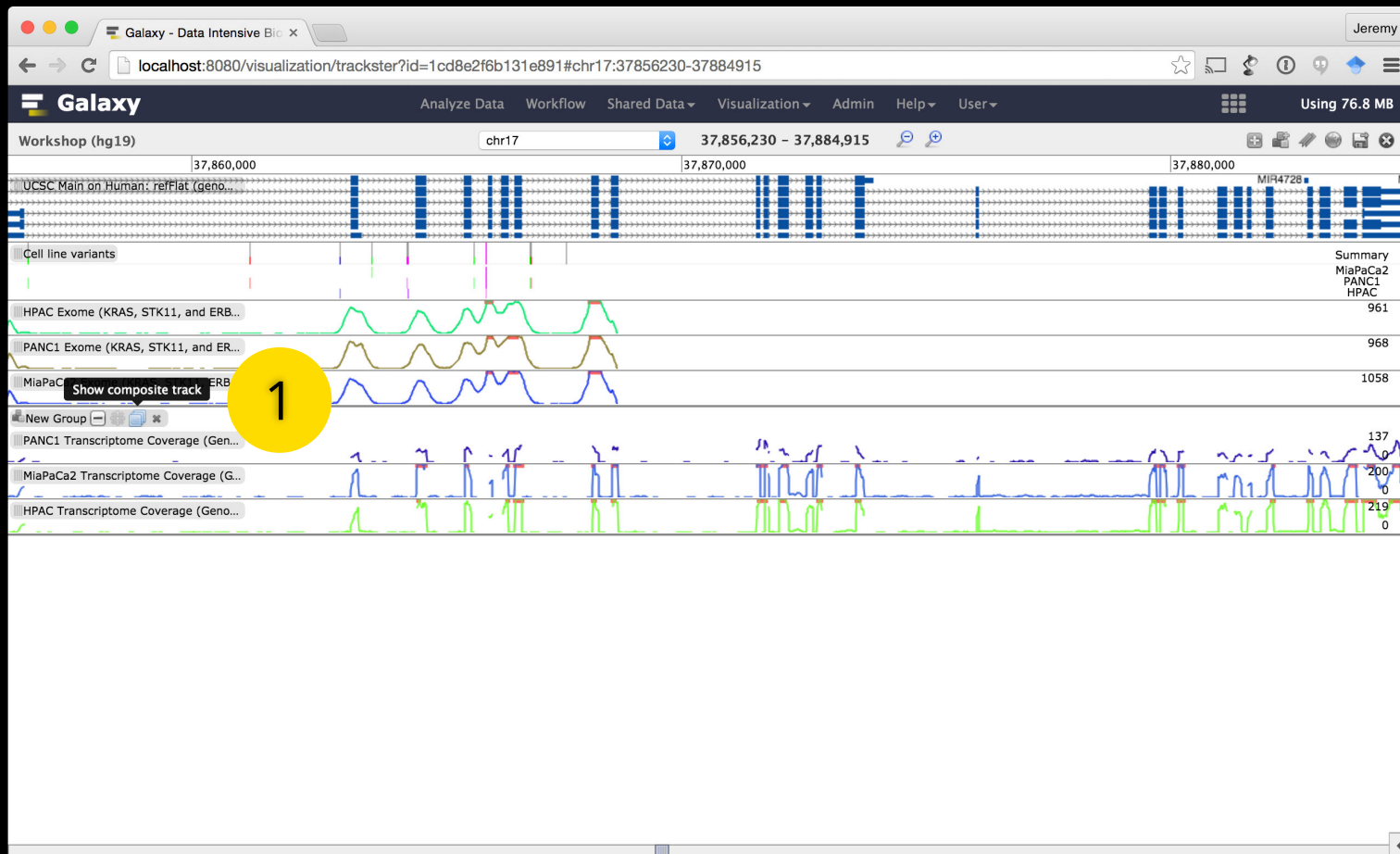
Back to Trackster: Rainbow Track for Coverage

4. Add transcriptome coverage tracks to group



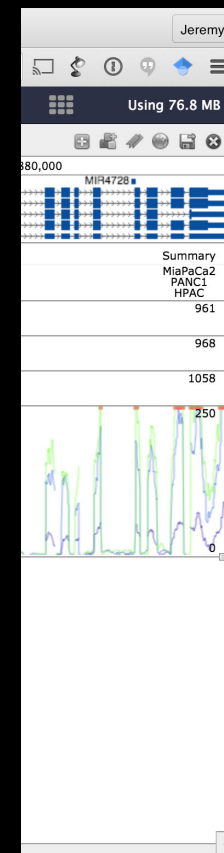
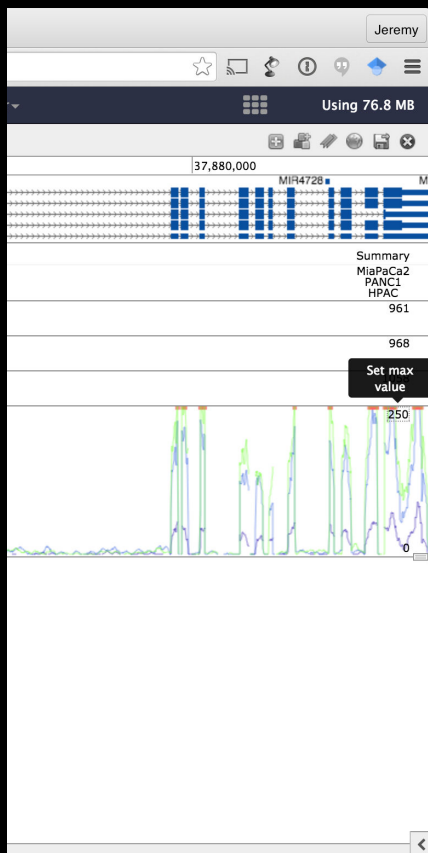
Back to Trackster: Rainbow Track for Coverage

5. Create composite track



Back to Trackster: Rainbow Track for Coverage

6. Adjust max and height; change name



Add More Data

Add RNA-seq mapped reads, variants, and assembled transcripts

Look at ERBB2

- ✦ bookmark

Look at STK11

- ✦ bookmark

Look at KRAS

- ✦ bookmark

Share and Publish

Galaxy

localhost:8080/visualization

Jeremy

Analyze DataWorkflowShared DataVisualizationAdminHelpUserUsing 76.8 MB

Saved Visualizations

search Q

[Advanced Search](#)

Create new visualization

☐	Title	Type	Dbkey	Tags	Sharing	Created	Last Updated↑
☐	Workshop	Trackster	hg19	0 Tags		~6 hours ago	~2 seconds ago
☐		Trackster	hg19	0 Tags		Jun 18, 2015	Jun 18, 2015
☐		Trackster	hg19	0 Tags		Mar 31, 2015	Mar 31, 2015

Open

Open in Circster

Edit Attributes

Copy

Share or Publish

Delete

1

Visualizations shared with you by others

No visualizations have been shared with you.

localhost:8080/visualization?f-sharing=All&sort=-update_time&f-tags=All&operation=Share+or+Publish&f-title=All&id=1cd8e2f6b13

Sharing and Publishing Vis x Galaxy x

localhost:8080/visualization/sharing?id=1cd8e2f6b131e891

Jeremy

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 76.8 MB

Share or Publish Visualization 'Workshop'

Make Visualization Accessible via Link and Publish It

This visualization is currently **accessible via link**.
Anyone can view and import this visualization by visiting the following URL:

<http://localhost:8080/u/jgoecks/v/workshop>

1

You can:

Disable Access to Visualization Link

Disables visualization's link so that it is not accessible.

Publish Visualization

Publishes the visualization to Galaxy's [Published Visualizations](#) section, where it is publicly listed and searchable.

Share Visualization with Individual Users

You have not shared this visualization with any users.

Share with a user

[Back to Visualizations List](#)

Share and Publish

Galaxy | Accessible Visuali... X

localhost:8080/u/jgoecks/v/workshop#chr17:37856230-37884915

Jeremy

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 76.8 MB

Accessible Visualization | Workshop

chr17

37,856,230 – 37,884,915

Import visualization

UCSC Main on Human: refFlat (geno...)

Cell line variants

HPAC Exome (KRAS, STK11, and ERB...)

PANC1 Exome (KRAS, STK11, and ER...)

MiaPaCa2 Exome (KRAS, STK11, ERB...)

New Group

37,860,000

37,870,000

37,880,000

MIR4728

MIE

Summary

MiaPaCa2

PANC1

HPAC

961

968

1058

250

0

About this Visualization

Author

jgoecks

Related items

All published items

Published items by jgoecks

Rating

Community

(0 ratings, 0.0 average)

Yours

Tags

Community: none

Yours:

Bookmarks

Jeremy

Demo: Visual Analysis

Topics

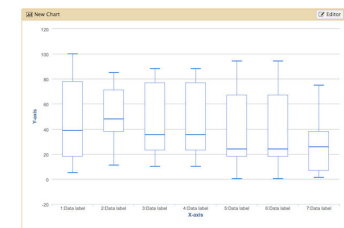
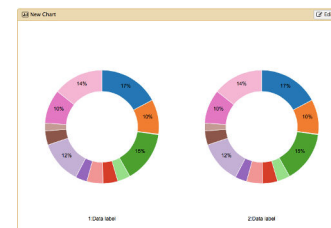
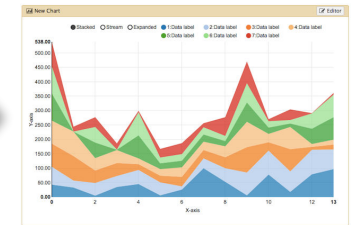
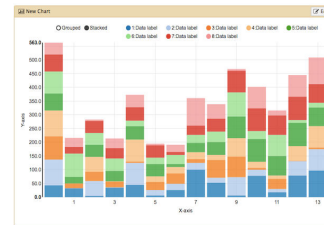
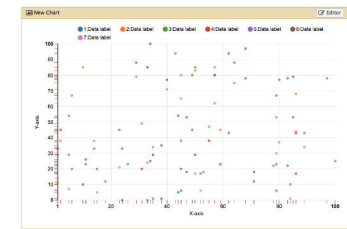
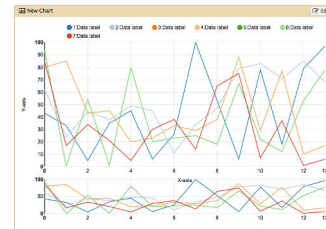
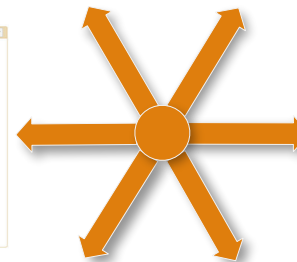
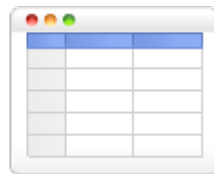
Visualization history and introduction

Numerical Visualizations

Biological Visualizations

Adding your own visualizations

What is Galaxy Charts?



Use Galaxy

Create Tabular Results

Visualize with Galaxy Charts

Import data files

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. An orange arrow labeled '1' points to the 'Shared Data' dropdown menu. The dropdown menu is open, showing options: 'Data Libraries', 'Data Libraries Beta', 'Published Histories', 'Published Workflows', 'Published Visualizations', and 'Published Pages'. An orange arrow labeled '2' points to the 'Data Libraries' option. Below the dropdown, the 'Data Library "Charts"' section is visible. An orange arrow labeled '3' points to the 'Name' column header. The table lists datasets with columns: 'Name', 'Data type', 'Date uploaded', and 'File size'. Two datasets are shown: 'amino_acid_features.txt' and 'http://www.compsysbio.org/bacteriome/dataset/functional_...'. An orange arrow labeled '4' points to the 'Go' button in the 'For selected datasets:' section.

Galaxy Analyze Data Workflow **Shared Data** Visualization Help User Using 2.1

Data Library "Charts"

Name

☐ amino_acid_features.txt

☐ http://www.compsysbio.org/bacteriome/dataset/functional_...

For selected datasets: Import to current history **Go**

Data type **Date uploaded** **File size**

amino_acid_features.txt	tabular	Mon Jun 30 04:13:31 2014 (UTC)	974 bytes
http://www.compsysbio.org/bacteriome/dataset/functional_...	tabular	Mon Jun 30 16:33:33 2014 (UTC)	81.6 KB

TIP: You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.

TIP: Several compression options are available for downloading multiple library datasets simultaneously:

- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats

Click on **Shared Data** and select **Data Libraries**. Navigate to the **Chart** library and import it into your history (*data reference: <http://dna.cs.byu.edu/treesaap> and bacteriome.org*).

Make a new chart (1 of 4)

1 → S1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt [eye] [pencil] [x]

3,989 lines
format: **tabular**, database: ?

uploaded tabular file

[save] [info] [refresh] [chart] 2 →

3 → Charts

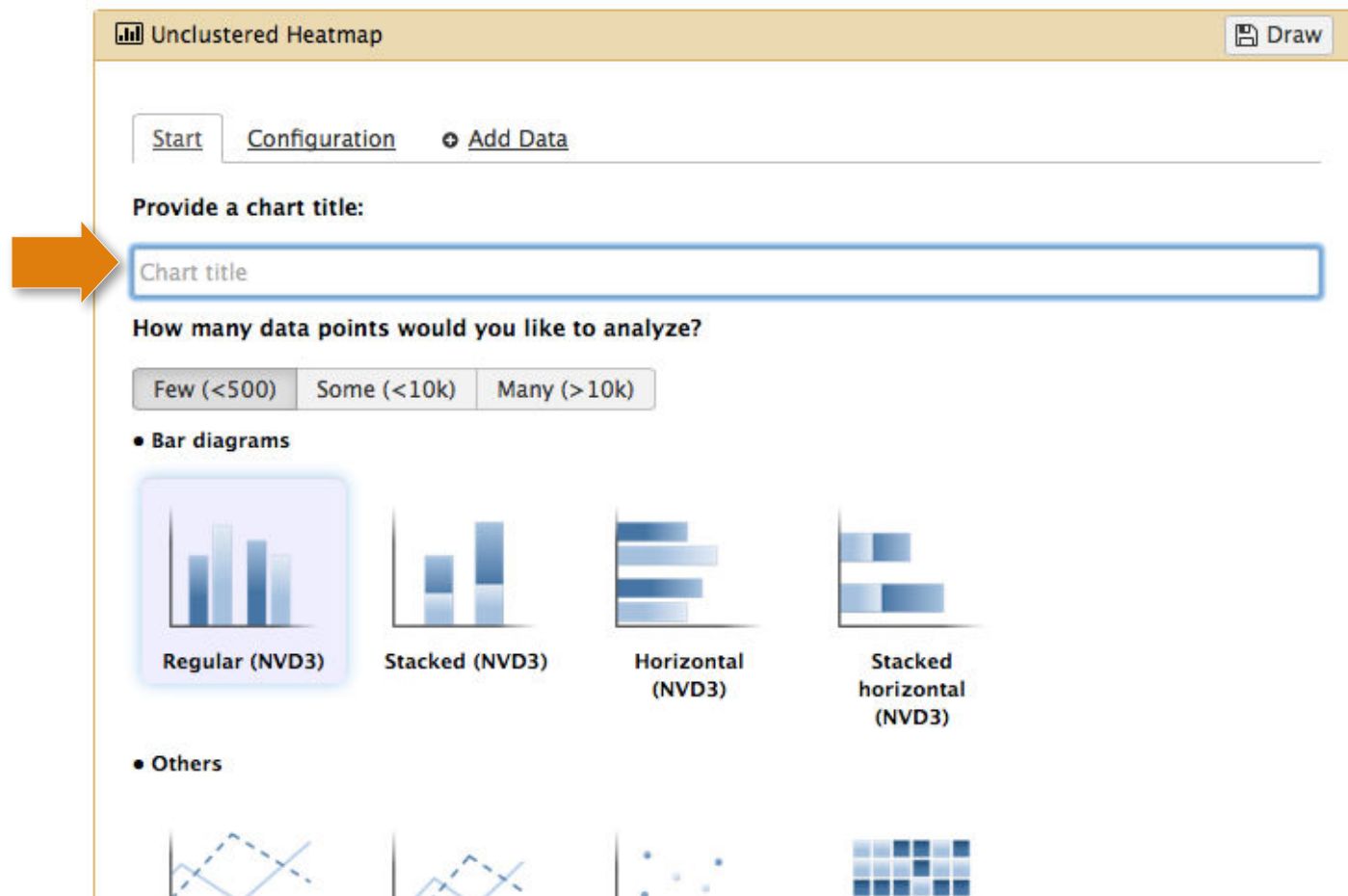
- Scatterplot
- Trackster

B4200 B4202 0.933934
B0779 B4058 0.933934
B0032 B0033 0.933183

The screenshot shows a web interface for uploading and visualizing data. An orange arrow labeled '1' points to the URL input field. Another orange arrow labeled '2' points to the visualization icon (a bar chart) in the toolbar. A third orange arrow labeled '3' points to the 'Charts' dropdown menu, which is open and shows options like 'Scatterplot' and 'Trackster'. Below the menu, a table of data is visible.

Wait for the upload to complete. Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

Few (<500) Some (<10k) Many (>10k)

• Bar diagrams

Regular (NVD3) Stacked (NVD3) Horizontal (NVD3) Stacked horizontal (NVD3)

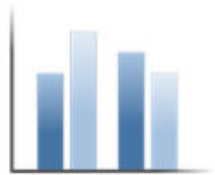
• Others

Name your chart **Unclustered Heatmap**.

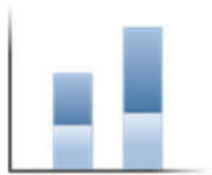
Select a chart type

Few (<500) Some (<10k) Many (>10k)

• Bar diagrams



Regular (NVD3)



Stacked (NVD3)

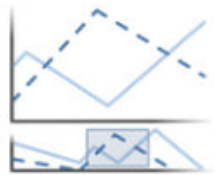


Horizontal
(NVD3)



Stacked
horizontal
(NVD3)

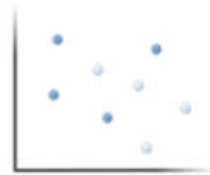
• Others



QLine with focus
(NVD3)



QLine chart
(NVD3)



QScatter plot
(NVD3)



QHeatmap
(Custom)

Double click on the **Heatmap** icon.

Select data columns

New Chart [Cancel] [Draw]

Start Configuration 1: Data label [Add Data]

Provide a label:

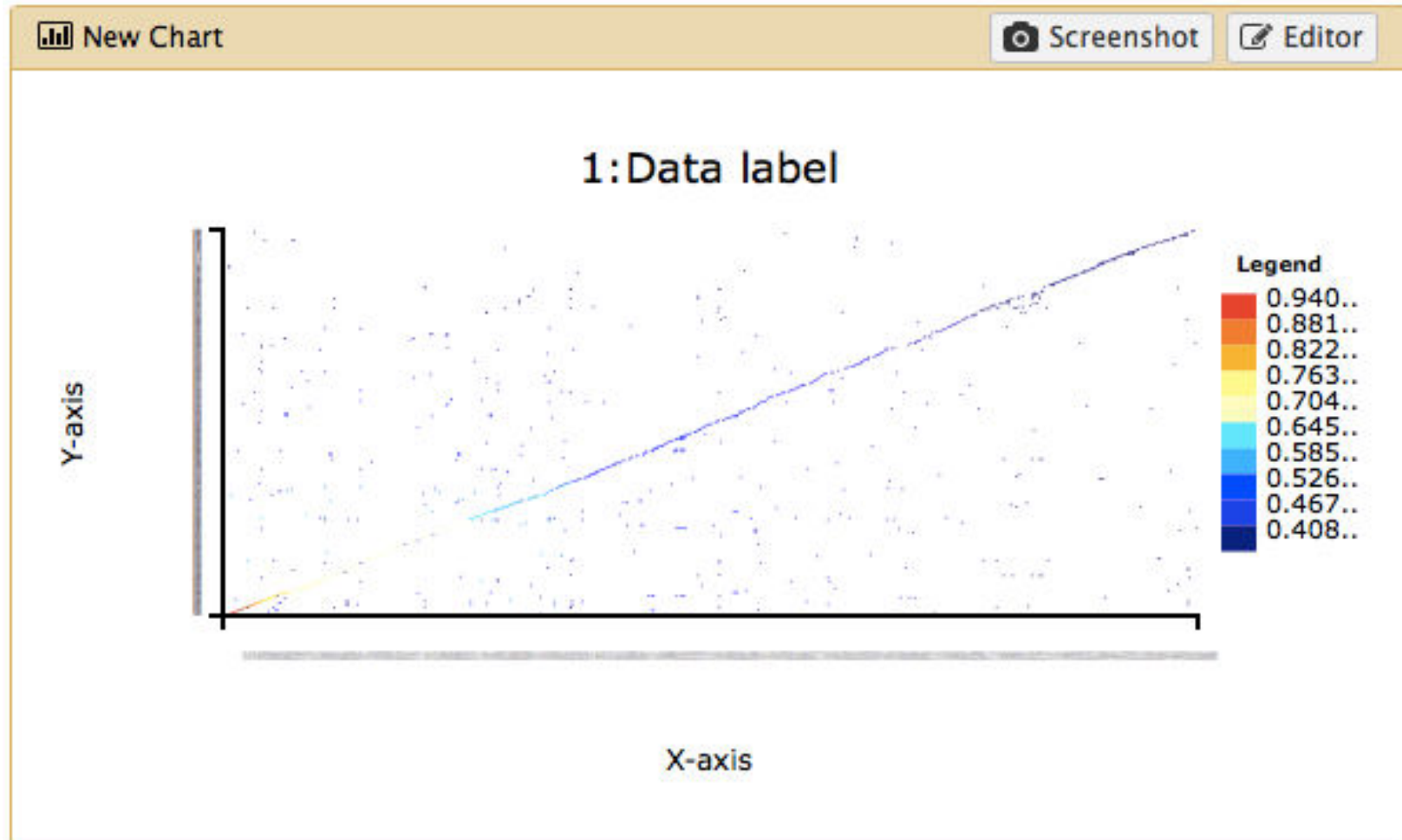
Data label

Select columns:

Column labels (all data labels)	Column: 1 [str]
Row labels (all data labels)	Column: 2 [str]
Observation	Column: 3 [float]

At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.

Unclustered Heatmap



Make a new chart (2 of 4)

1 → S1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt [eye] [pencil] [x]

3,989 lines
format: **tabular**, database: ?

uploaded tabular file

[save] [info] [refresh] [bar chart] 2 →

3 → Charts

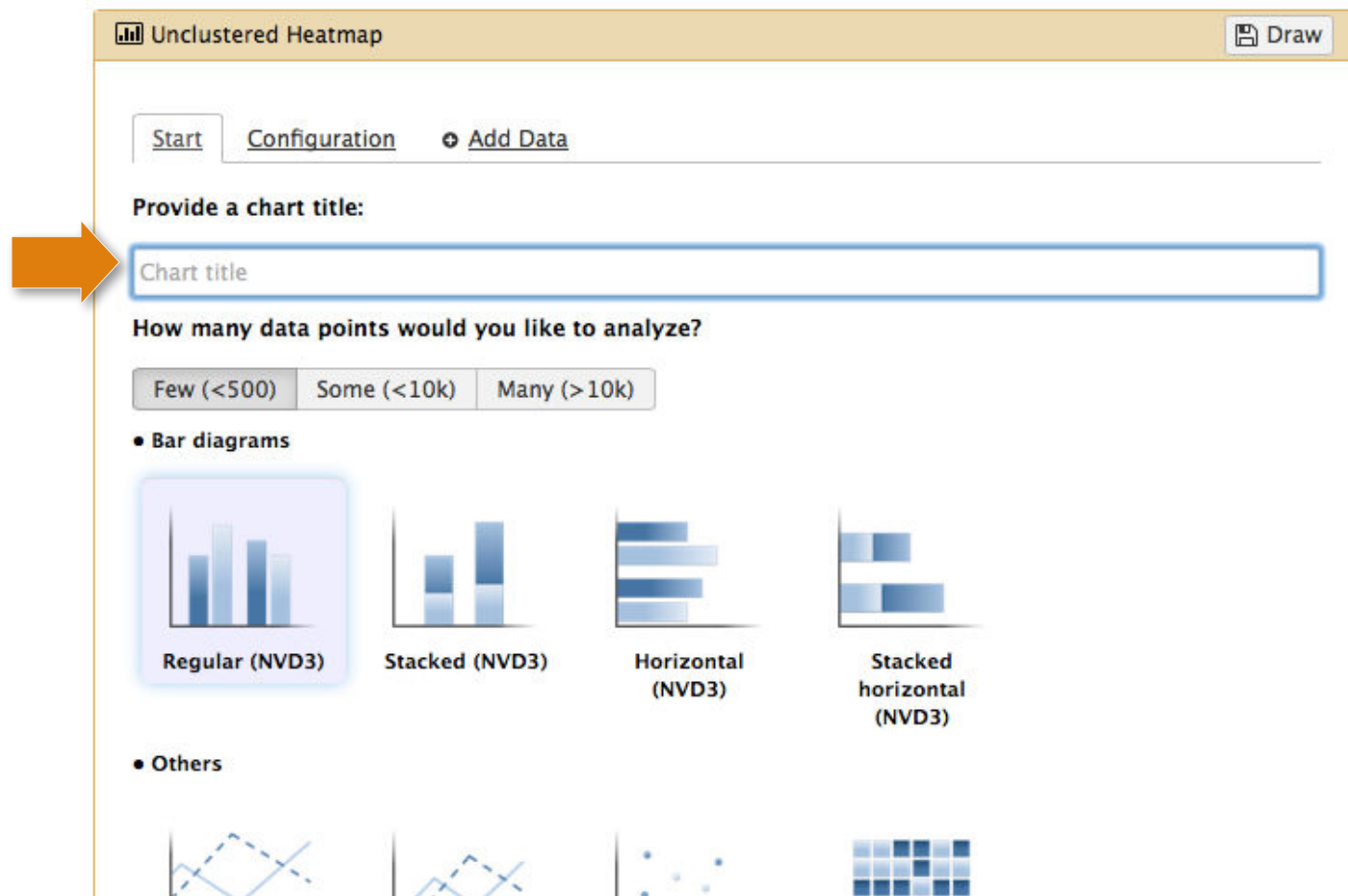
- Scatterplot
- Trackster

B4200	B4202	0.933934
B0779	B4058	0.933934
B0032	B0033	0.933183

The screenshot shows a web interface for a dataset. An orange arrow labeled '1' points to the dataset URL. Below the URL, it shows the number of lines (3,989) and the format (tabular). An orange arrow labeled '2' points to the visualization icon (a bar chart) in the toolbar. An orange arrow labeled '3' points to the 'Charts' menu item in the dropdown. The dropdown also shows 'Scatterplot' and 'Trackster'. Below the menu, a table of data is visible.

Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

Few (<500) Some (<10k) Many (>10k)

• Bar diagrams

Regular (NVD3) Stacked (NVD3) Horizontal (NVD3) Stacked horizontal (NVD3)

• Others

Name your chart **Clustered Heatmap**.

Select a new chart type

- Area charts



QRegular (NVD3)



QExpanded (NVD3)



QStream (NVD3)



Pie chart (NVD3)

- Data processing (requires 'charts' tool from Toolshed)



Histogram (NVD3)



QDiscrete Histogram (jqPlot)



QBox plot (jqPlot)



QClustered Heatmap (Custom)



Double click on the **Clustered Heatmap** icon.

Select data columns

New Chart [Cancel] [Draw]

Start Configuration 1: Data label [Add Data]

Provide a label:

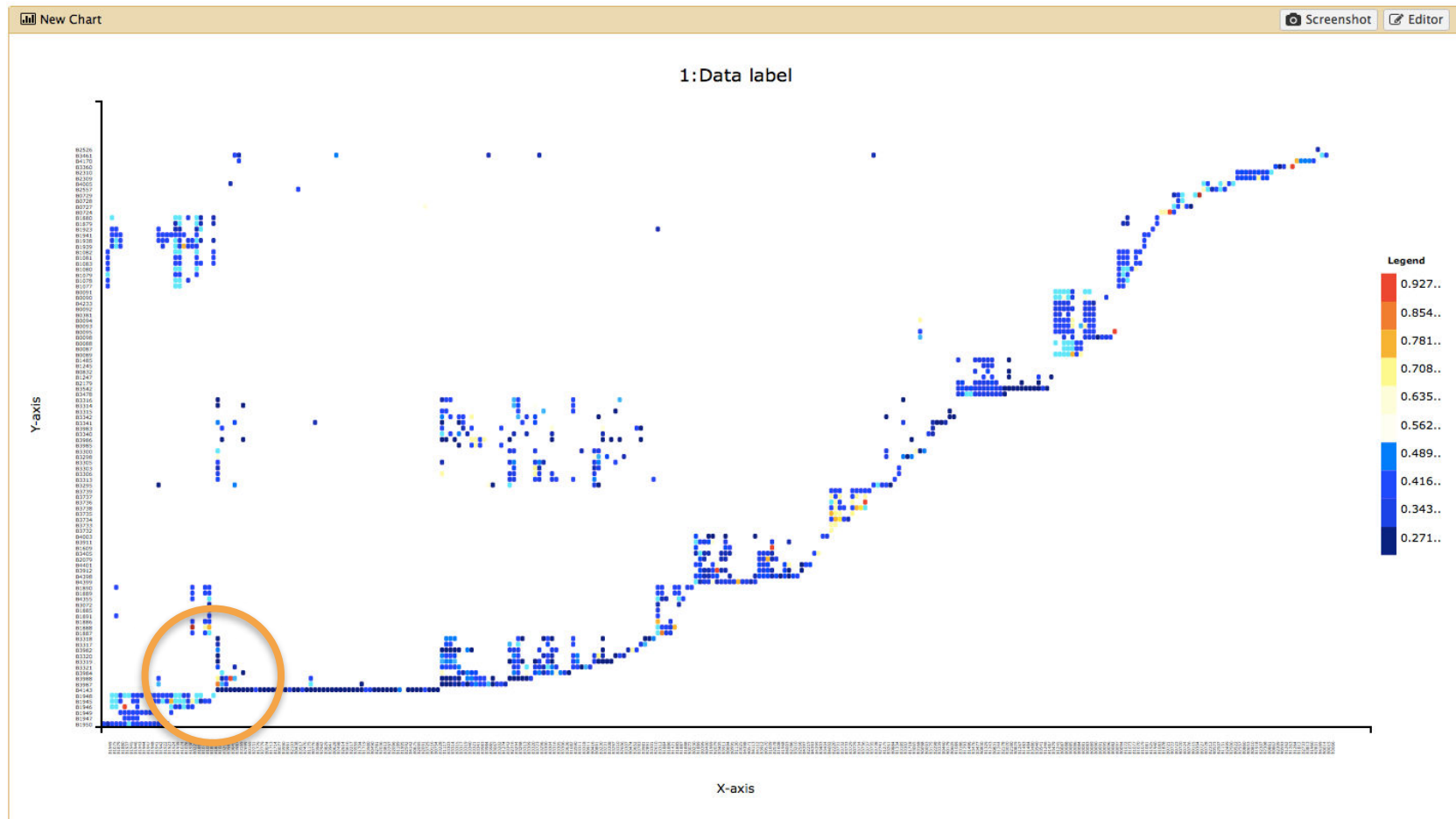
Data label

Select columns:

Column labels (all data labels)	Column: 1 [str]
Row labels (all data labels)	Column: 2 [str]
Observation	Column: 3 [float]

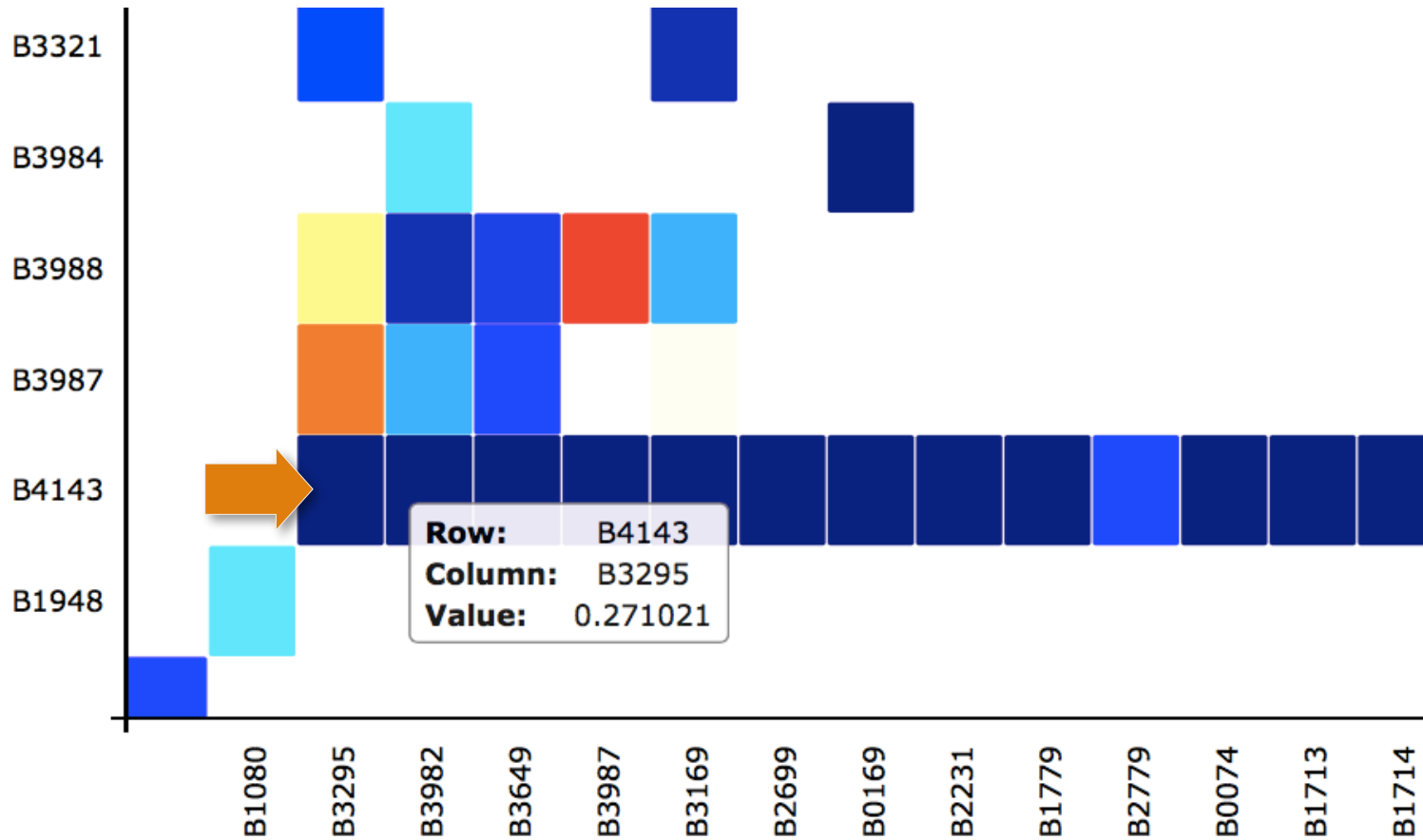
At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.

Clustered Heatmap



Use the **mouse wheel** or **your touch pad** to zoom into the **highlighted area**.

Enlarged view



Tooltips popup if you move the mouse pointer over a box. Here the interaction between **B4143** and **B3295** is highlighted. Click on **Editor** again to further customize this chart.

Chart configuration

 New Chart



Start

Configuration

1: Data label 

 Add Data

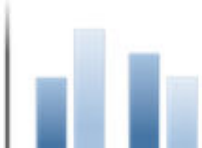
Provide a chart title:

New Chart

How many data points would you like to analyze?

Few (<500)Some (<10k)Many (>10k)

• Bar diagrams



Go to the **Configuration** tab.

Chart settings

X axis:

Axis label

X-axis

Provide a label for the axis.

Axis value type

Auto

Select the value type of the axis.

Y axis:

Axis label

Y-axis

Provide a label for the axis.

Axis value type

Auto

Select the value type of the axis.

Others:

Show legend

Yes

No

Would you like to add a legend?

Color scheme

Jet

Select a color scheme for your heatmap

Url template

http://someurl.com?id=__LABEL__

Enter a url to link the labels with external sources. Use __LABEL__ as placeholder.



Heatmap specific options are **highlighted**. Feel free to set **axis labels** or other options.

Define a URL template

Yes

No

Would you like to add a legend?

Jet

Select a color scheme for your heatmap

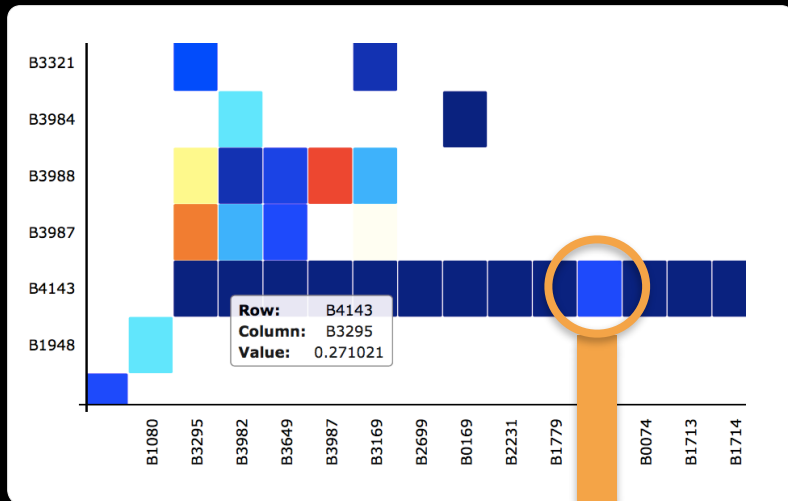


http://www.ncbi.nlm.nih.gov/geoprofiles/?term=__LABEL__

Enter a url to link the labels with external sources. Use __LABEL__ as placeholder

Paste a **database URL** into the template URL field and add the **__LABEL__** tag. You may use **<http://www.ncbi.nlm.nih.gov>** or any other database. Click on **Draw** to redraw the chart.

Data points linked to web sources



Double click on a **box** and the browser will open two new tabs using the previously defined **URL template**.

NCBI Resources How To Sign in to NCBI

GEO Profiles

GEO Profiles b4143 Search Save search Advanced Help

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Subgroup effect Send to: Filters: Manage Filters

Results: 1 to 20 of 47

1. [groEL - Stress factor RpoS regulon in exponential-phase bacteria](#)

Annotation: groEL, molecular chaperone GroEL (multiple annotations exist)
Organism: Escherichia coli K-12
Reporter: GPL199, mopA_b4143_at (ID_REF), GDS3123, 1037522 (Gene ID), 913705 (Gene ID), 948665 (Gene ID), 959980 (Gene ID), b4143 (ORF)
DataSet type: Expression profiling by array, transformed count, 6 samples
ID: 49311248

[GEO DataSets](#) [Gene](#) [Profile neighbors](#) [Chromosome neighbors](#)

2. [groL - Indole-3-acetic acid effect on Escherichia coli](#)

Annotation: groL, Cpn60 chaperonin GroEL, large subunit of GroESL (multiple annotations exist)
Organism: Escherichia coli, Escherichia coli K-12
Reporter: GPL189, 1240 (ID_REF), GDS2181, b4143 (ORF)
DataSet type: Expression profiling by array, count, 12 samples
ID: 27346540

[GEO DataSets](#) [Gene](#) [Profile neighbors](#) [Chromosome neighbors](#)

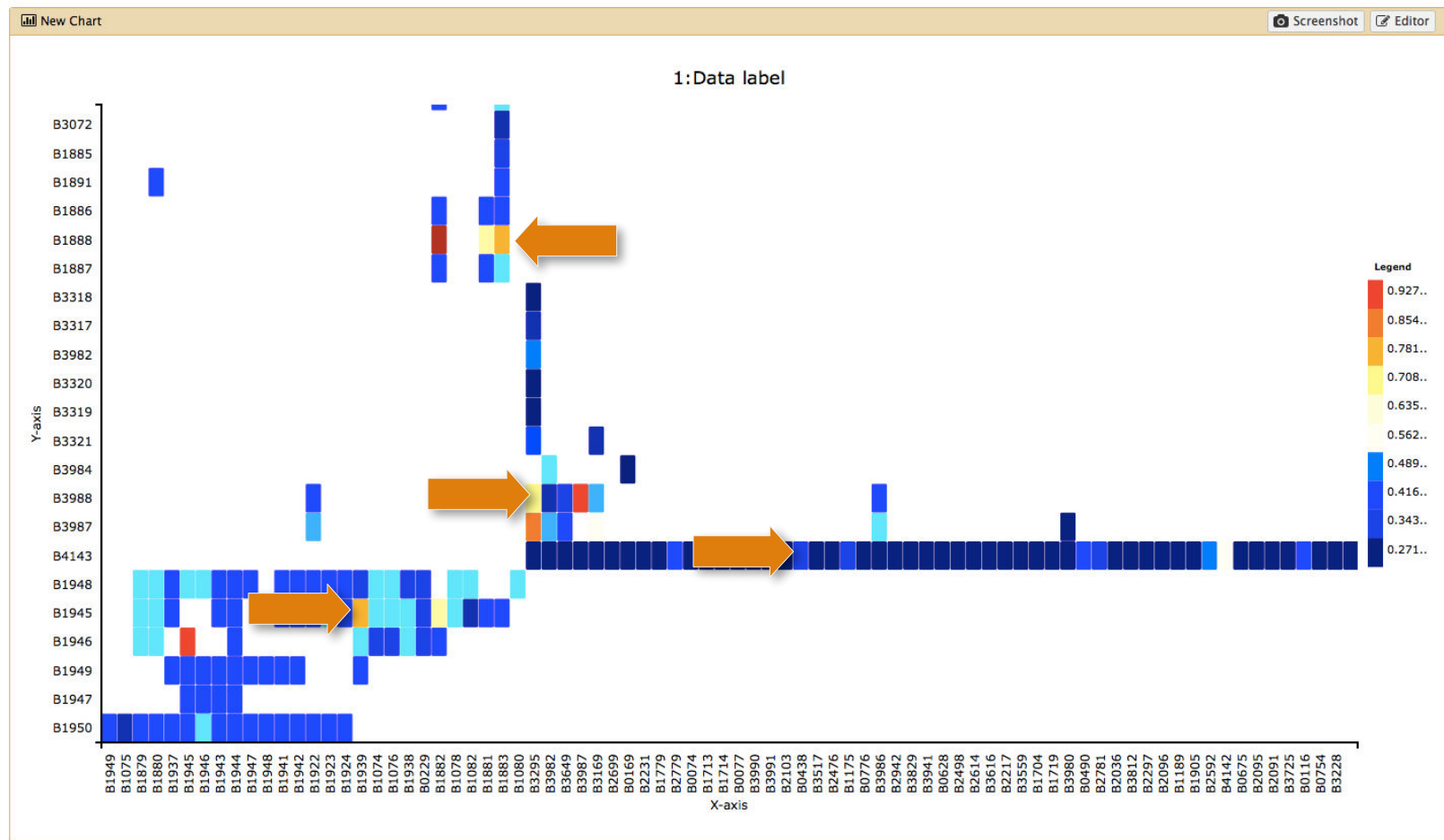
Profile data
Download profile data

Profile pathways
Find pathways

Find related data
Database: Select
Find items

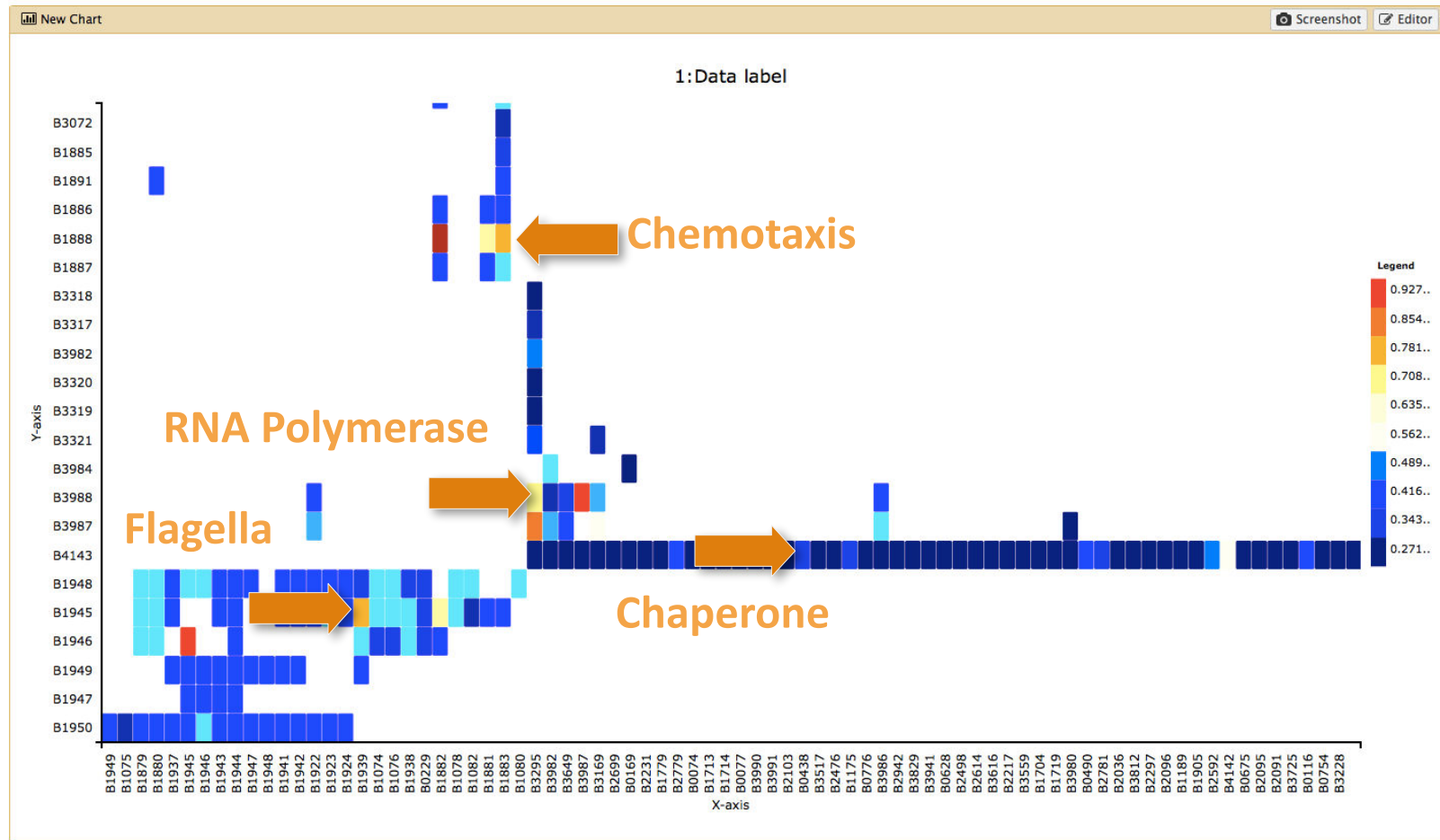
Search details
b4143[All Fields]

Cluster selection and analysis



Select one element from each **highlighted row**. What are the corresponding **protein functions**?

Identified protein categories



Please return to the **Editor**.

Make a new chart (3 of 4)

1 → S1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt [eye] [pencil] [x]

3,989 lines
format: **tabular**, database: ?

uploaded tabular file

[save] [info] [refresh] [bar chart] 2 →

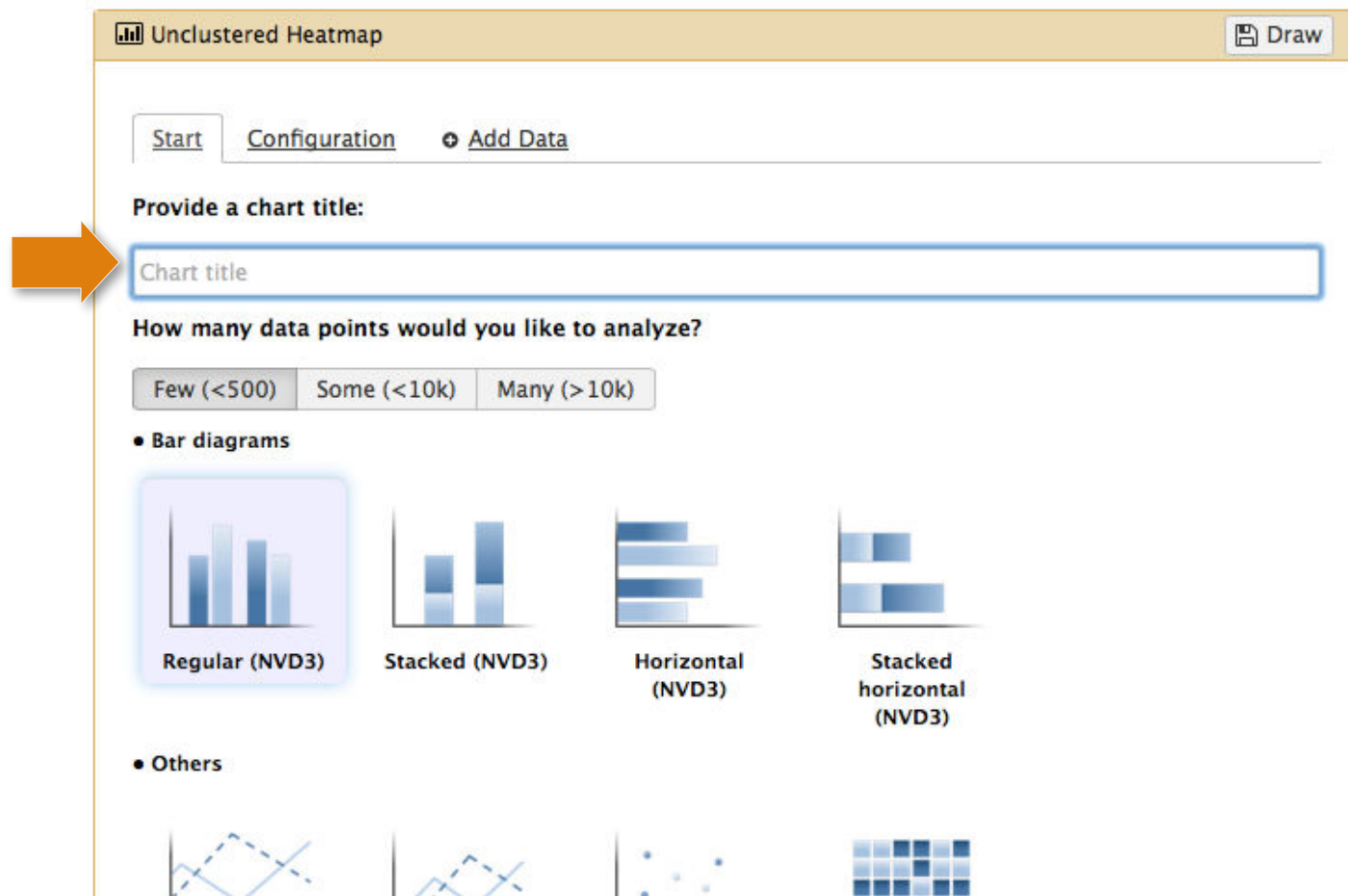
3 → Charts

- Scatterplot
- Trackster

B4200 B4202 0.933934
B0779 B4058 0.933934
B0032 B0033 0.933183

Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

☐ Few (<500) ☐ Some (<10k) ☐ Many (>10k)

• Bar diagrams

☒ Regular (NVD3) ☐ Stacked (NVD3) ☐ Horizontal (NVD3) ☐ Stacked horizontal (NVD3)

• Others

☐ Line ☐ Area ☐ Scatter ☐ Heatmap

Name your chart **Score Histogram**.

Analyze the score distribution

- Area charts



QRegular (NVD3)



QExpanded (NVD3)



QStream (NVD3)



Pie chart (NVD3)

- Data processing (requires 'charts' tool from Toolshed)



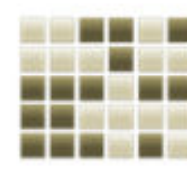
Histogram (NVD3)



QDiscrete Histogram (jqPlot)



QBox plot (jqPlot)



QClustered Heatmap (Custom)

Double click on the **Histogram** icon and click on **Draw**.

Give your chart a name

 Unclustered Heatmap   Draw

Please select data columns before drawing the chart.

[Start](#) [Configuration](#)

1: Data label 

[+ Add Data](#)

Provide a label:

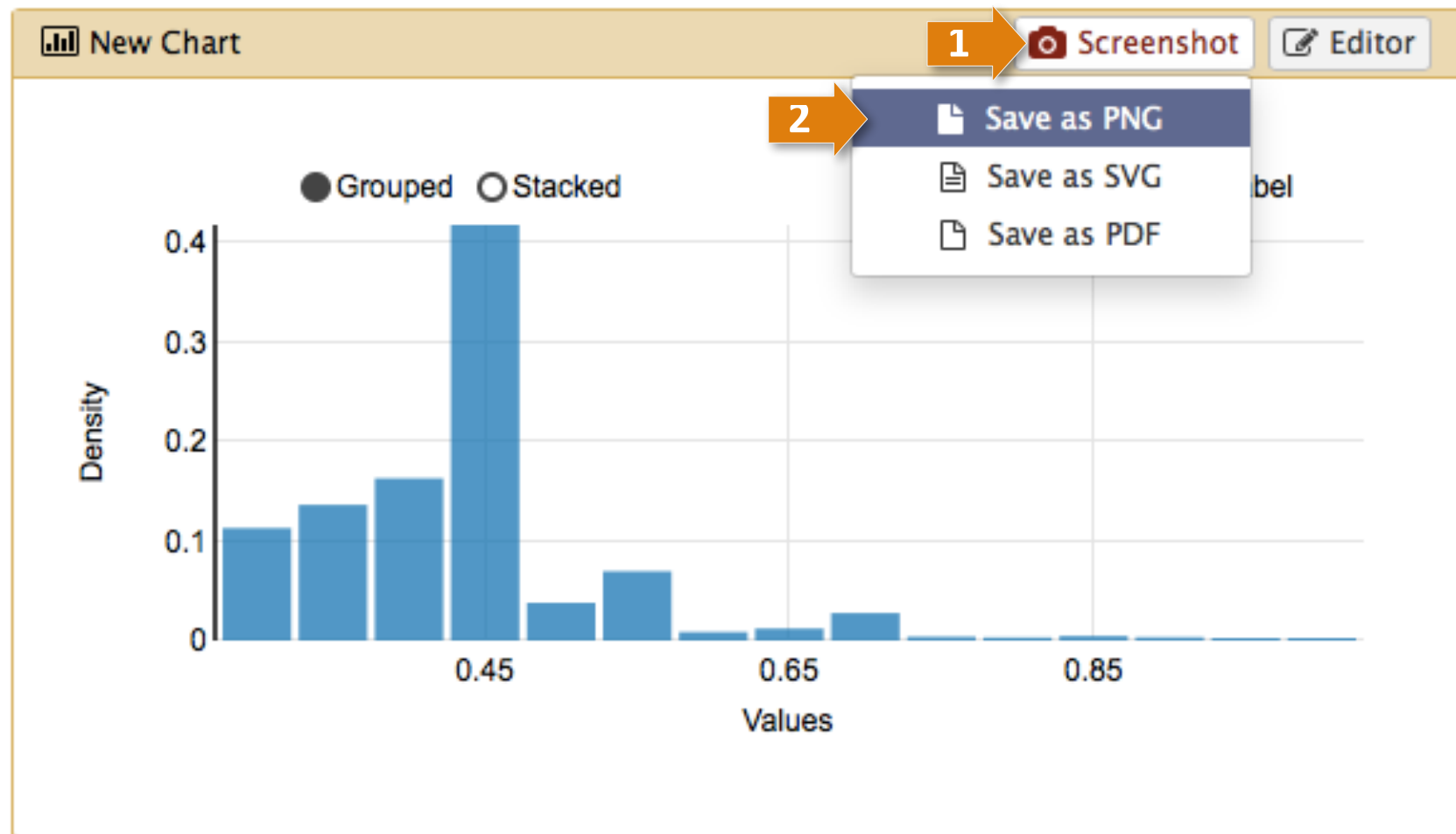
Select columns:

Observations



Click on **Draw**.

Export as PNG



Click on **Screenshot** and select **Save as PNG**. Finally, return to the **Editor** again.

Make a new chart (4 of 4)

1 → S1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt [eye] [pencil] [x]

3,989 lines
format: **tabular**, database: ?

uploaded tabular file

[save] [info] [refresh] [bar chart] 2 →

3 → Charts

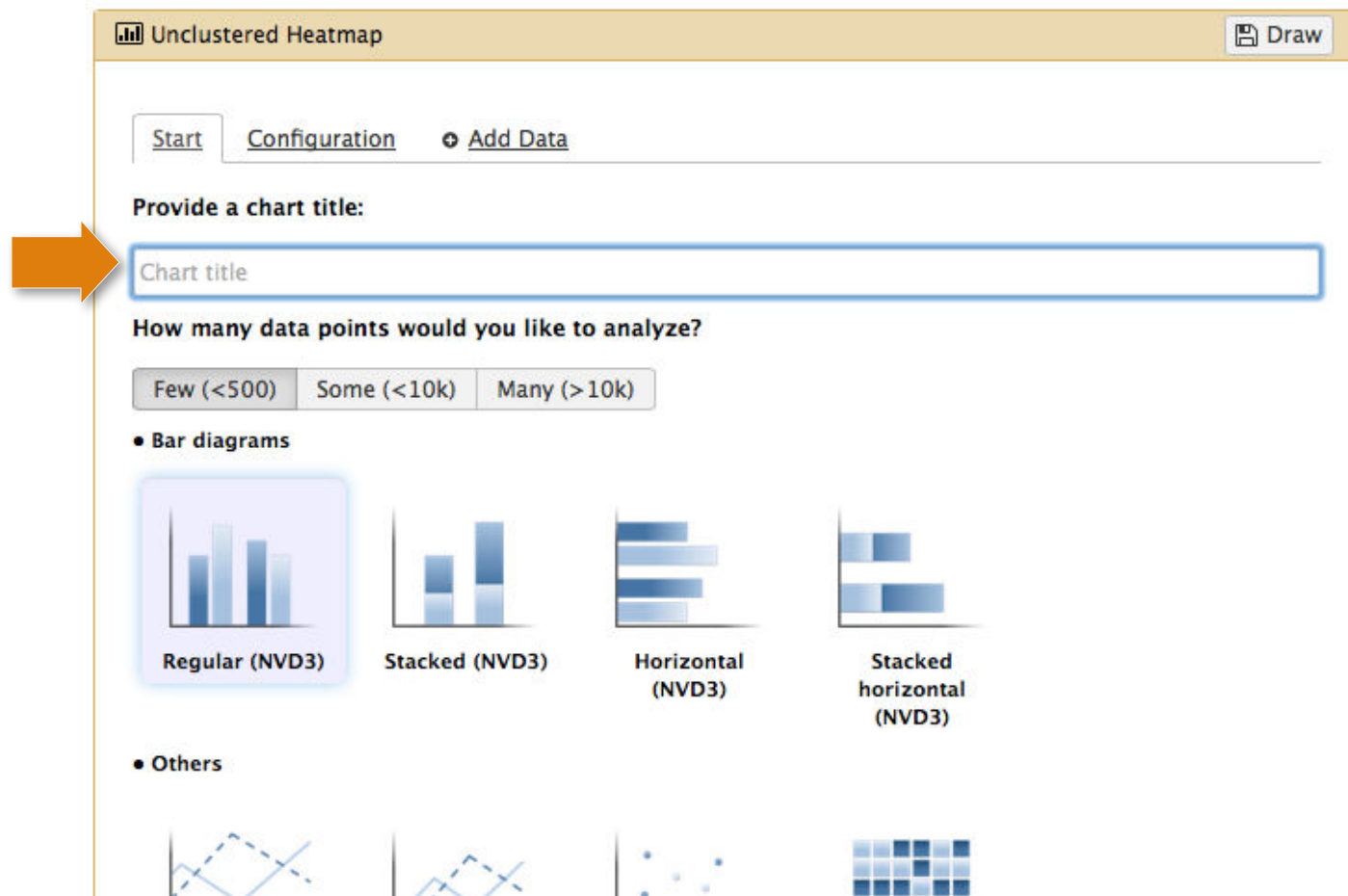
- Scatterplot
- Trackster

B4200	B4202	0.933934
B0779	B4058	0.933934
B0032	B0033	0.933183

The screenshot shows a web interface for a dataset. An orange arrow labeled '1' points to the dataset URL. Below the URL, it shows the number of lines (3,989) and the format (tabular). An orange arrow labeled '2' points to a bar chart icon in a toolbar. An orange arrow labeled '3' points to a dropdown menu that is open, showing 'Charts' as the selected option, with 'Scatterplot' and 'Trackster' as other options. Below the menu, a table of data is visible.

Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



Unclustered Heatmap Draw

[Start](#) [Configuration](#) [Add Data](#)

Provide a chart title:

Chart title

How many data points would you like to analyze?

☐ Few (<500) ☐ Some (<10k) ☐ Many (>10k)

• Bar diagrams

☒ Regular (NVD3) ☐ Stacked (NVD3) ☐ Horizontal (NVD3) ☐ Stacked horizontal (NVD3)

• Others

☐ Line ☐ Area ☐ Scatter ☐ Heatmap

Name your chart **Discrete Histogram**.

Analyze the protein distribution

- Area charts



ⓀRegular (NVD3)



ⓀExpanded (NVD3)



ⓀStream (NVD3)



Pie chart (NVD3)

- Data processing (requires 'charts' tool from Toolshed)



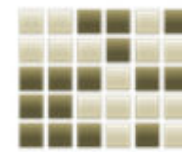
Histogram (NVD3)



ⓀDiscrete Histogram (jqPlot)



ⓀBox plot (jqPlot)



ⓀClustered Heatmap (Custom)

Double click on the **Discrete Histogram** icon.

Add more data

 Unclustered Heatmap  Draw

[Start](#) [Configuration](#) [1: Data label](#)  [+ Add Data](#) 

Provide a label:

Select columns:

Observations



Click on **Add Data**.

Select a second data group

The screenshot shows a 'New Chart' dialog box with a yellow header bar. In the top right corner, there are 'Cancel' and 'Draw' buttons. An orange arrow with the number '2' points to the 'Draw' button. Below the header, there are tabs: 'Start', 'Configuration', '1: Data label', and '2: Data label'. The '2: Data label' tab is selected. To the right of the tabs is a '+ Add Data' button. Below the tabs, the text 'Provide a label:' is followed by a text input field containing 'Data label'. Below that, the text 'Select columns:' is followed by a list box. An orange arrow with the number '1' points to the 'Observations' item in the list box. The list box shows 'Observations' selected, and below it, a dropdown menu shows 'Column: 2 [str]'.

New Chart

Cancel Draw

Start Configuration 1: Data label 2: Data label + Add Data

Provide a label:

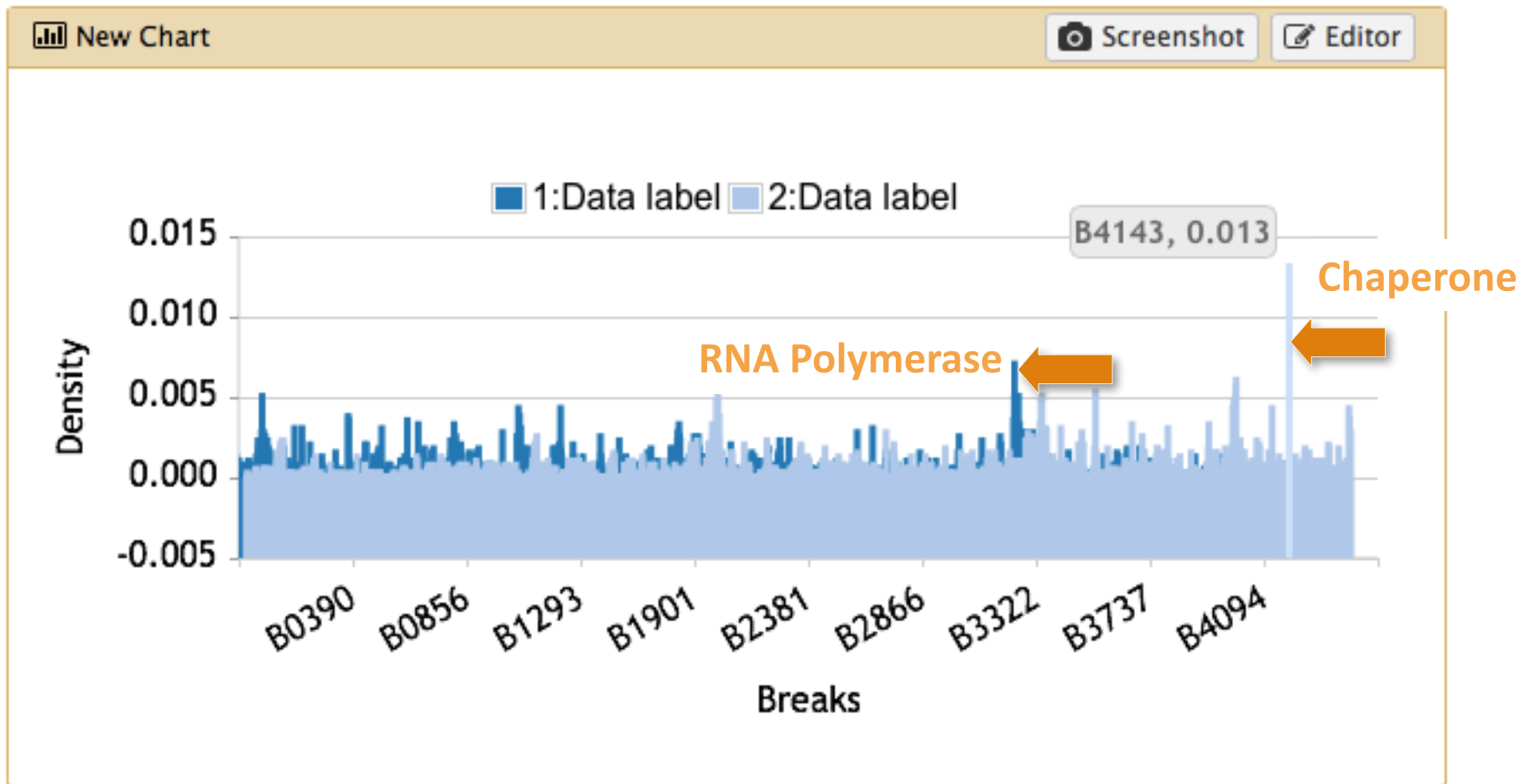
Data label

Select columns:

1 Observations Column: 2 [str]

At first click on **Observations** and select **Column 2**. Then, click on **Draw**.

Which proteins have most interactions?



Done with Part I.

Scratchbook

Activate the Scratchbook

The screenshot shows the Galaxy web interface. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Help, and User. On the right side of the top bar, there is a grid icon (Scratchbook) and a status indicator 'Using 2.1 MB'. An orange arrow points to the grid icon. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: Get Data, Lift-Over, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, Operate on Genomic Intervals, Statistics, Graph/Display Data, Regional Variation, Multivariate Analysis, Evolution, Motif Tools, Multiple Alignments, and FASTA manipulation. The main content area displays 'Welcome to Galaxy on the Cloud' managed by CloudMan. The right sidebar shows the 'History' section with a list of recent jobs. The top job is '25: Charts on data 1' and the second is '24: amino acid features.txt'. Below these, there is a job titled '1: http://www.compsysbio.org/bacteriome/dataset/functiona... interactions.txt' which is a tabular file with 3,989 lines. A table preview is shown below the job details.

1	2	3
B1882	B1888	1.000000
B0728	B0729	0.966967
B1812	B3360	0.956456
B4200	B4202	0.933934
B0779	B4058	0.933934
B0022	B0022	0.023182

Activate the **Scratchbook** by clicking on the above icon.

Activate the Scratchbook

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'Visualization' menu is open, showing 'New Track Browser' and 'Saved Visualizations'. An orange arrow points to 'Saved Visualizations'. The left sidebar lists various tools under categories like 'Get Data', 'Text Manipulation', 'Filter and Sort', etc. The main area displays 'Welcome to Galaxy on the Cloud' managed by CloudMan. The right sidebar shows a 'History' panel with a list of recent jobs, including '25: Charts on data 1', '24: amino acid features.txt', and '1: http://www.compsysbio.org/bacteriome/dataset/functiona... interactions.txt'. The bottom of the history panel shows a table with columns 1, 2, and 3, containing genomic data.

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

Get Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Regional Variation
Multivariate Analysis
Evolution
Motif Tools
Multiple Alignments
FASTA manipulation

Welcome to Galaxy on the Cloud
managed by CloudMan

History

Unnamed history
2.1 MB

25: Charts on data 1

24: amino acid features.txt

1: <http://www.compsysbio.org/bacteriome/dataset/functiona... interactions.txt>

3,989 lines
format: tabular, database: ?

uploaded tabular file

1	2	3
B1882	B1888	1.000000
B0728	B0729	0.966967
B1812	B3360	0.956456
B4200	B4202	0.933934
B0779	B4058	0.933934
B0022	B0022	0.022182

Click on **Saved Visualizations**.

Activate the Scratchbook

The screenshot displays the Galaxy web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'Saved Visualizations' panel is open, showing a search bar and a table of saved visualizations. An orange arrow points to the 'Unclustered Heatmap' visualization. The background shows a 'History' panel with a list of recent jobs, including '25: Charts on data 1' and '24: amino acid features.txt'.

Saved Visualizations

search

[Advanced Search](#)

☐ **Title** **Type** **Dbkey** **Tags** **Sharing** **Created** **Last Updated ↑**

☐ Unclustered Heatmap Charts [0 Tags](#) [Accessible, Published](#) ~4 hours ago ~3 hours ago

☐ Clustered Heatmap Charts [0 Tags](#) [Accessible, Published](#) ~21 hours ago ~5 hours ago

For 0 selected items:

Visualizations shared with you by others

No visualizations have been shared with you.

History

Unnamed history
2.1 MB

25: Charts on data 1

24: amino acid features.txt

1: <http://www.compsysbio.org/bacteriome/dataset/functiona>
[l_interactions.txt](#)

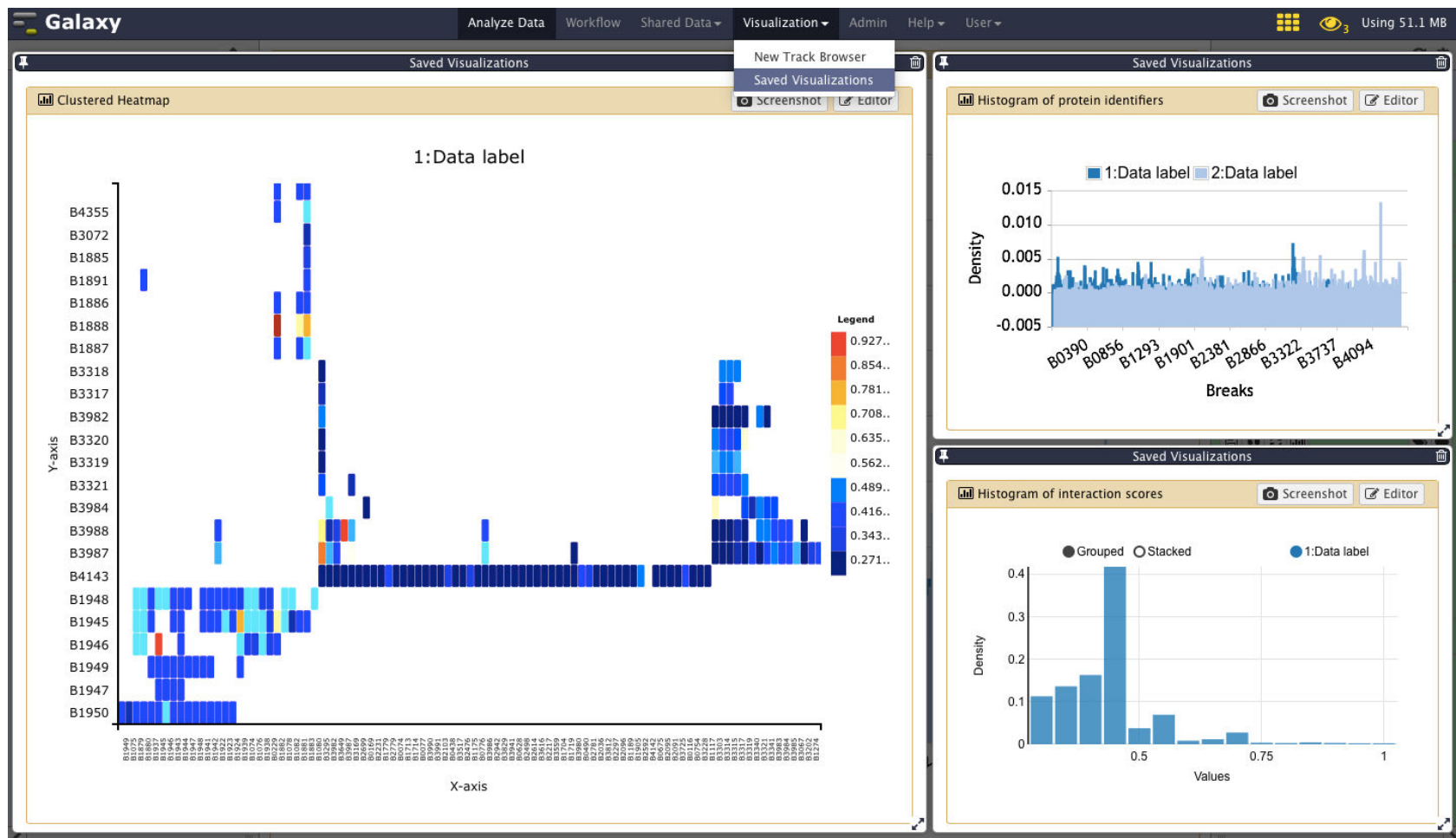
3,989 lines
format: tabular, database: ?

uploaded tabular file

1	2	3
B1882	B1888	1.000000
B0728	B0729	0.966967
B1812	B3360	0.956456
B4200	B4202	0.933934
B0779	B4058	0.933934
B0022	B0022	0.022182

Select a Visualization and repeat the process by selecting **Saved Visualizations** again.

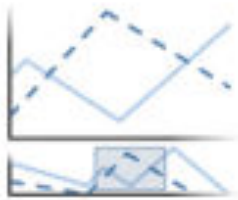
Scratchbook for multiple charts



Resize all visualizations so they fit into the screen.

More Examples

Create a pie chart



QLine with focus
(NVD3)



QLine chart
(NVD3)



QScatter plot
(NVD3)



QHeatmap
(Custom)

• Area charts



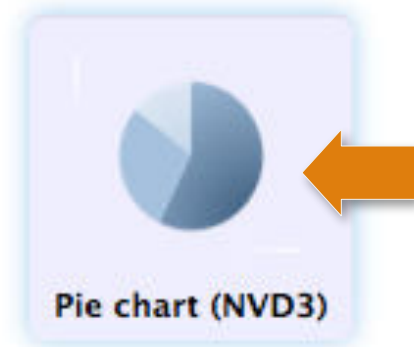
QRegular (NVD3)



QExpanded
(NVD3)



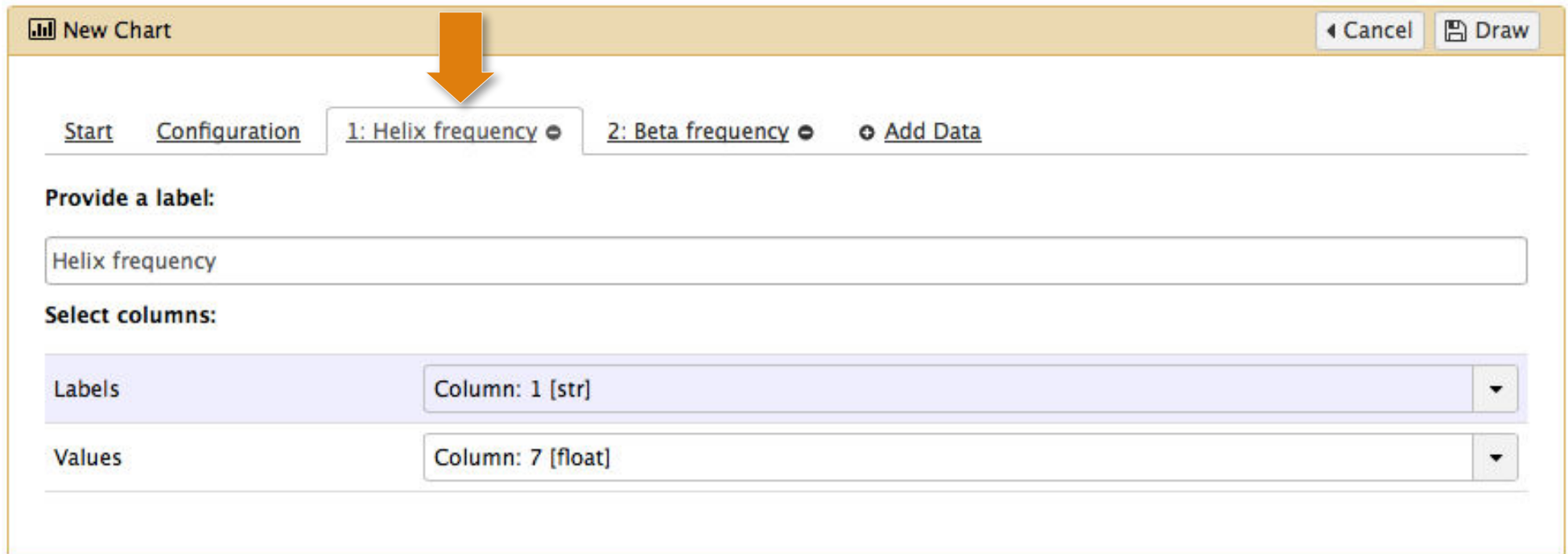
QStream (NVD3)



Pie chart (NVD3)

Select the imported datasets, create a new chart and select **Pie chart**. Then, click on **Add data**.

Add first data group



The screenshot shows a 'New Chart' dialog box with a yellow header bar. The header contains a chart icon, the text 'New Chart', and two buttons: 'Cancel' and 'Draw'. Below the header, there are three tabs: 'Start', 'Configuration', and '1: Helix frequency'. An orange arrow points to the '1: Helix frequency' tab. To the right of the tabs are two more tabs: '2: Beta frequency' and 'Add Data'. Below the tabs, the text 'Provide a label:' is followed by a text input field containing 'Helix frequency'. Below that, the text 'Select columns:' is followed by two rows. The first row has a label 'Labels' and a dropdown menu showing 'Column: 1 [str]'. The second row has a label 'Values' and a dropdown menu showing 'Column: 7 [float]'.

New Chart

Cancel Draw

Start Configuration 1: Helix frequency 2: Beta frequency Add Data

Provide a label:

Helix frequency

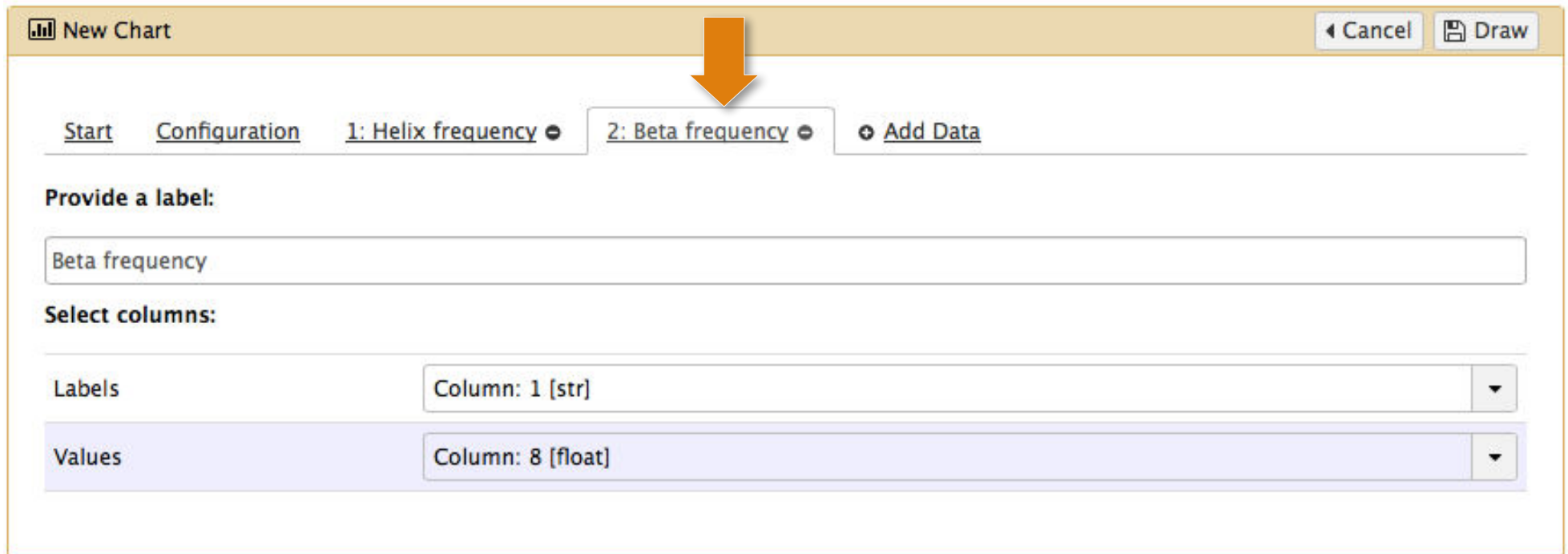
Select columns:

Labels Column: 1 [str]

Values Column: 7 [float]

Configure the **Helix frequency** column.

Add second data group



The screenshot shows a 'New Chart' dialog box with a yellow title bar. In the top right corner are 'Cancel' and 'Draw' buttons. Below the title bar is a tabbed interface with three tabs: 'Start', 'Configuration', and '1: Helix frequency'. The '2: Beta frequency' tab is selected and highlighted with an orange arrow pointing down to it. To the right of the tabs is an 'Add Data' button. Below the tabs, the 'Provide a label:' section has a text input field containing 'Beta frequency'. The 'Select columns:' section contains two rows: 'Labels' with a dropdown menu showing 'Column: 1 [str]' and 'Values' with a dropdown menu showing 'Column: 8 [float]'. The 'Values' row is highlighted with a light blue background.

New Chart

Cancel Draw

Start Configuration 1: Helix frequency 2: Beta frequency Add Data

Provide a label:

Beta frequency

Select columns:

Labels Column: 1 [str]

Values Column: 8 [float]

Configure the **Beta frequency** column.

Configure the pie chart

New Chart Cancel Draw

Start Configuration 1: Helix frequency 2: Beta frequency Add Data

Pie chart settings:

Donut ratio 50%
Determine how large the donut hole will be.

Show legend Yes No
Would you like to add a legend?

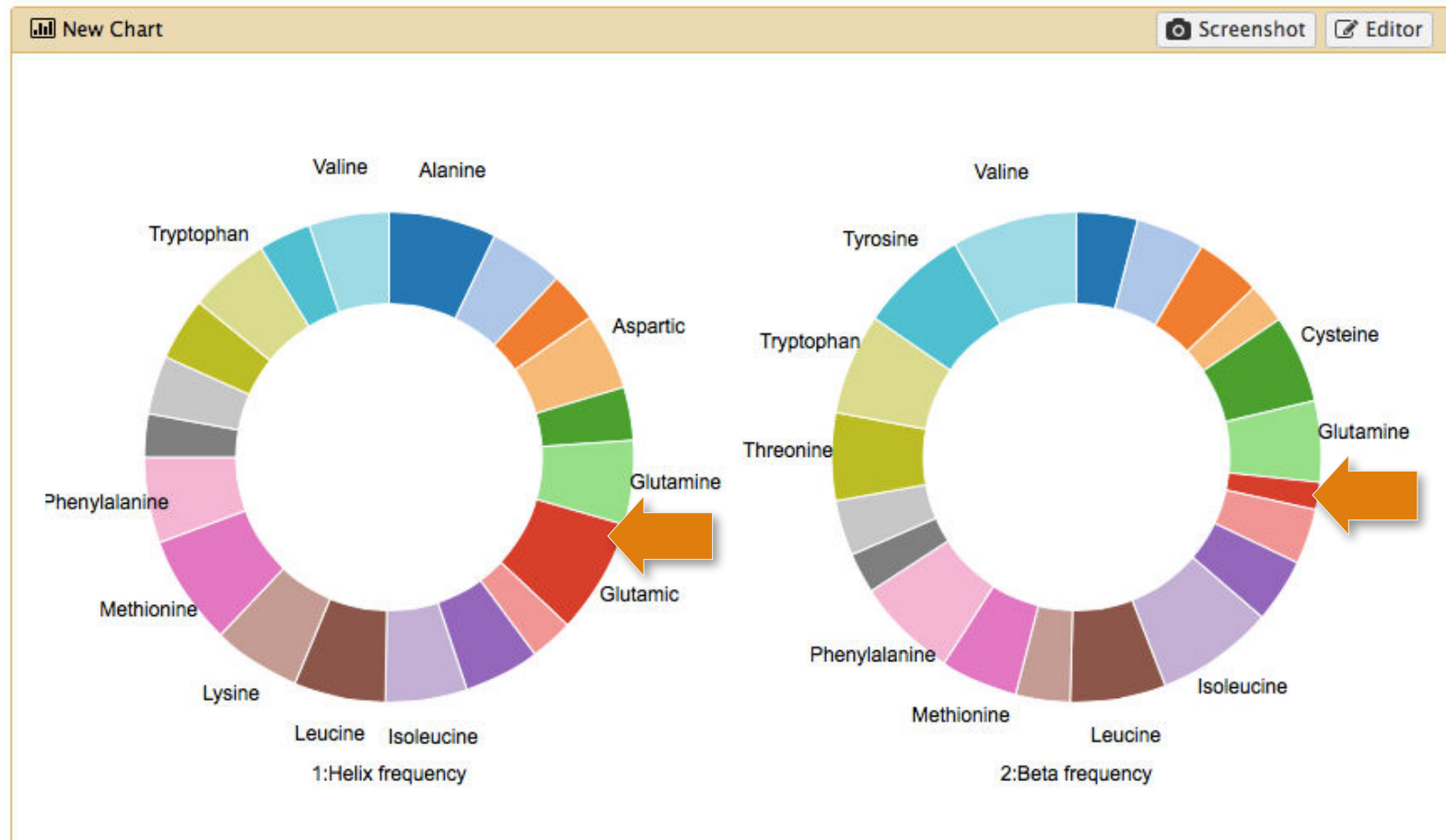
Label settings:

Donut label Label column
What would you like to show for each slice?

Show outside Yes No
Would you like to show labels outside the donut?

Configure the **Pie chart** as shown above. Then, click on **Draw**.

Configure the pie chart



Glutamic acids seem to fit much better into **helices** than **beta sheets**. In other words, "Aspartic and Glutamic Acids are Important for Alpha-helix Folding", JBSD 2007.

Create a bar diagram

Amino Acid Features (Bar diagram) Cancel Draw

[Start](#) [Configuration](#) [1: Hydrophobicity](#) [2: Membrane](#) [3: Flexibility](#) [4: Helix](#) [5: Beta](#) [Add Data](#)

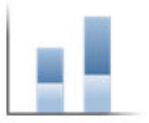
Provide a chart title:

Amino Acid Features (Bar diagram)

How many data points would you like to analyze?

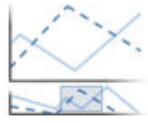
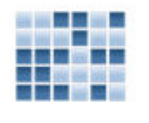
☒ Few (<500) ☐ Some (<10k) ☐ Many (>10k)

• Bar diagrams

Regular (NVD3) Stacked (NVD3) Horizontal (NVD3) Stacked horizontal (NVD3)

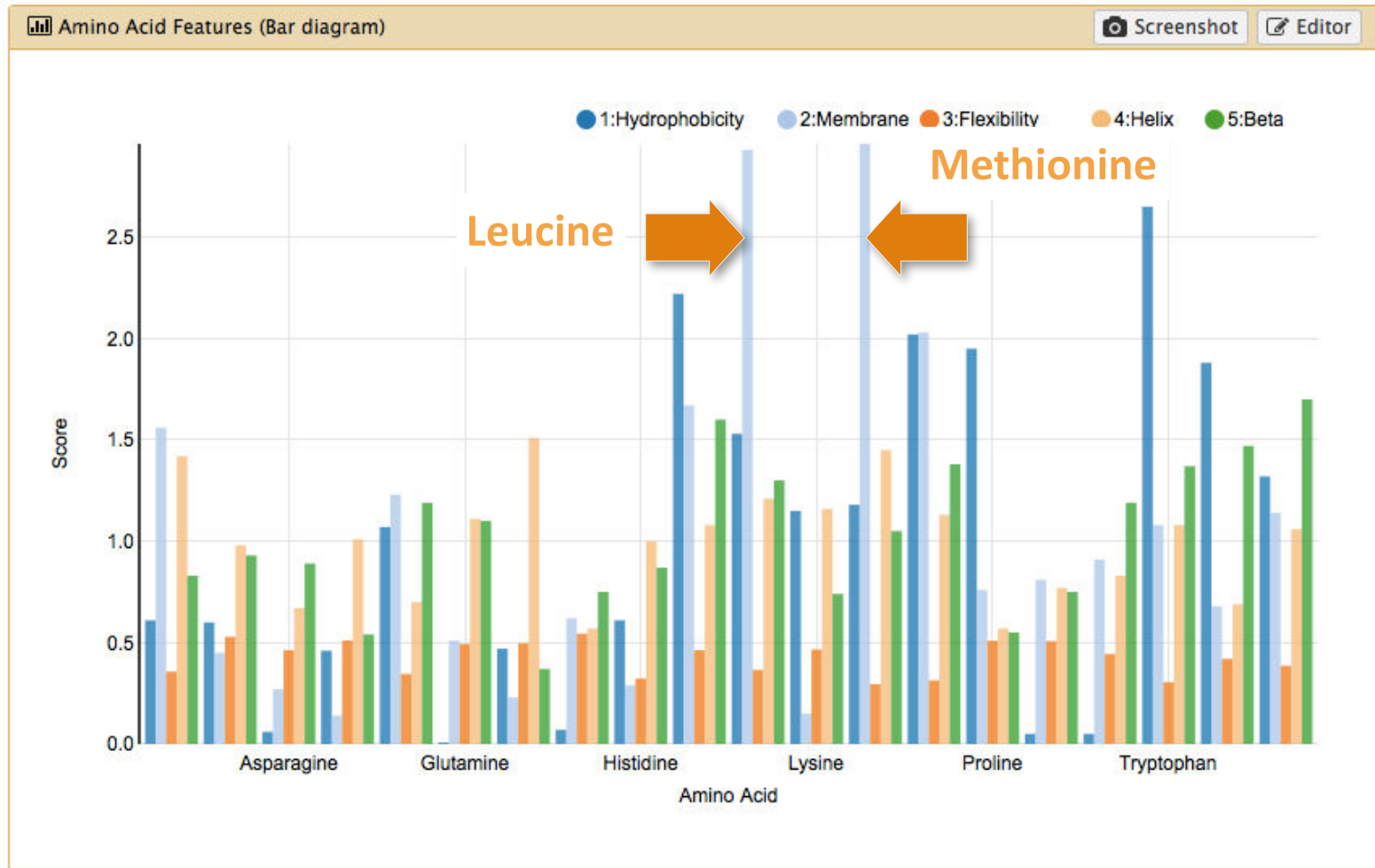
• Others

QLine with focus QLine chart QScatter plot QHeatmap

Create data groups for the following features: **Hydrophobicity**, **Membrane frequency**, **Flexibility**, **Helix frequency** and **Beta frequency**.

Bar diagram of amino acid features



Use the **tooltips** to identify the amino acids which are likely to be found within membrane proteins.

Topics

Visualization history and introduction

Numerical Visualizations

Biological Visualizations

Adding your own visualizations

Adding your own Visualizations

Go to **config/plugins/visualizations/charts**

charts/others/YOURVIZNAME

Add three files to this directory:

Logo (logo.png)

Configuration (config.js)

Wrapper (wrapper.js)

charts/types.js

Rebuild by typing 'npm install' and 'grunt'

Workshop Materials

Will be available on training day page:
<http://gcc2015.tsl.ac.uk/training-day/>

For this workshop:

- ✦ Galaxy page on usegalaxy.org with:
- ✦ published history
- ✦ published visualization

<https://usegalaxy.org/u/jeremy/p/visualization-workshop>