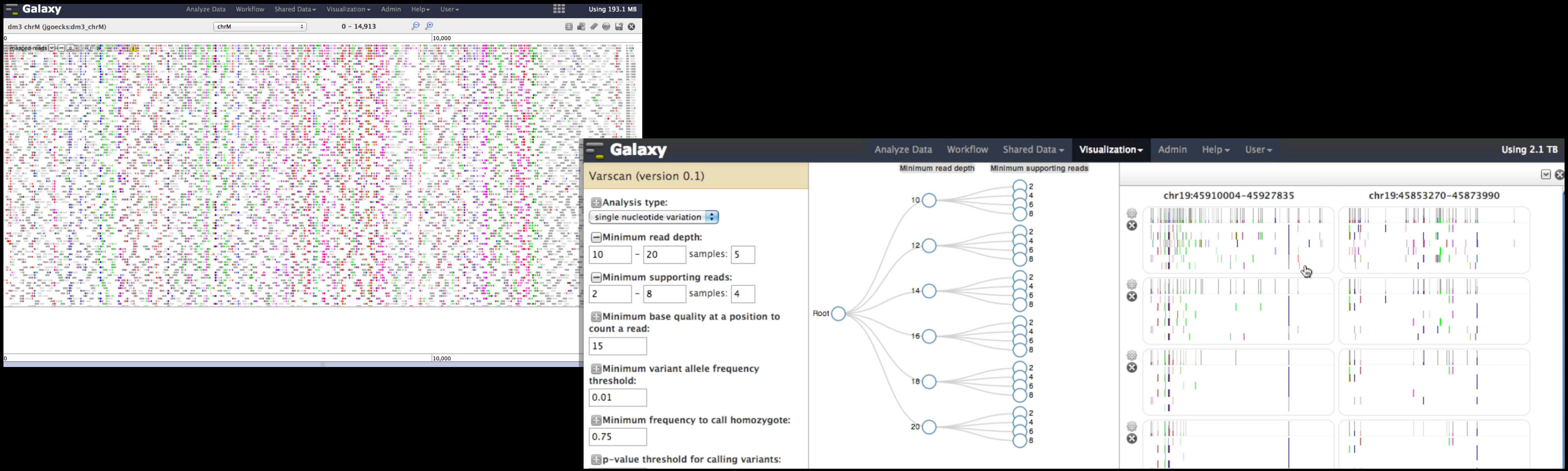


Varscan (version 0.1)

Analysis type: single nucleotide variation

Minimum read depth: 10 - 20 samples: 5

Minimum supporting reads: 2 - 8 samples: 4

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:

Galaxy Community Conference 2015: Visualization Workshop

Aysam Guerler and
Carl Eberhard

Jeremy Goecks

JOHNS HOPKINS
UNIVERSITY

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

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?

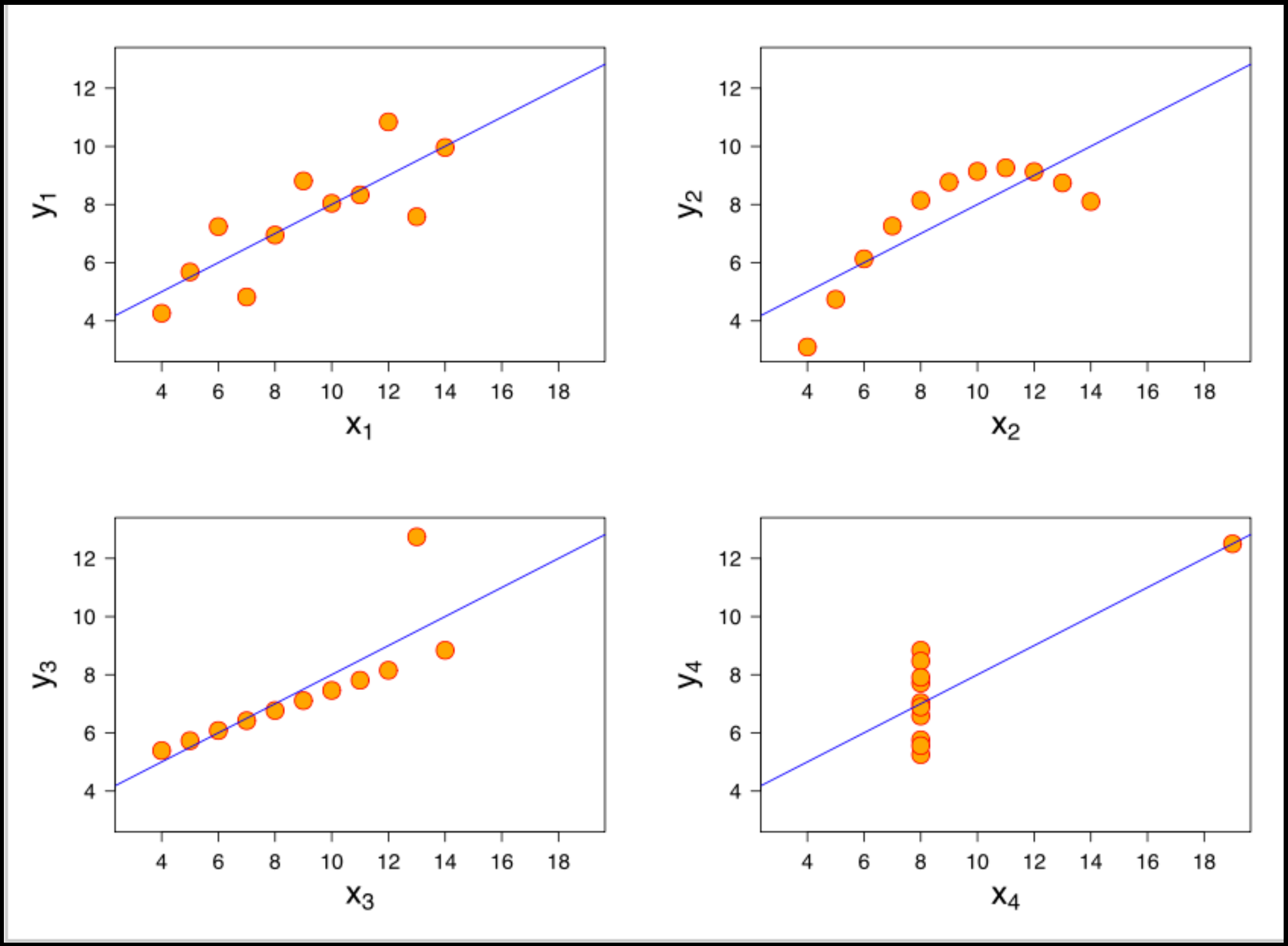
Why Visualize?

Quick check: did it work?

Exploration and hypothesis generation

Sharing/publishing

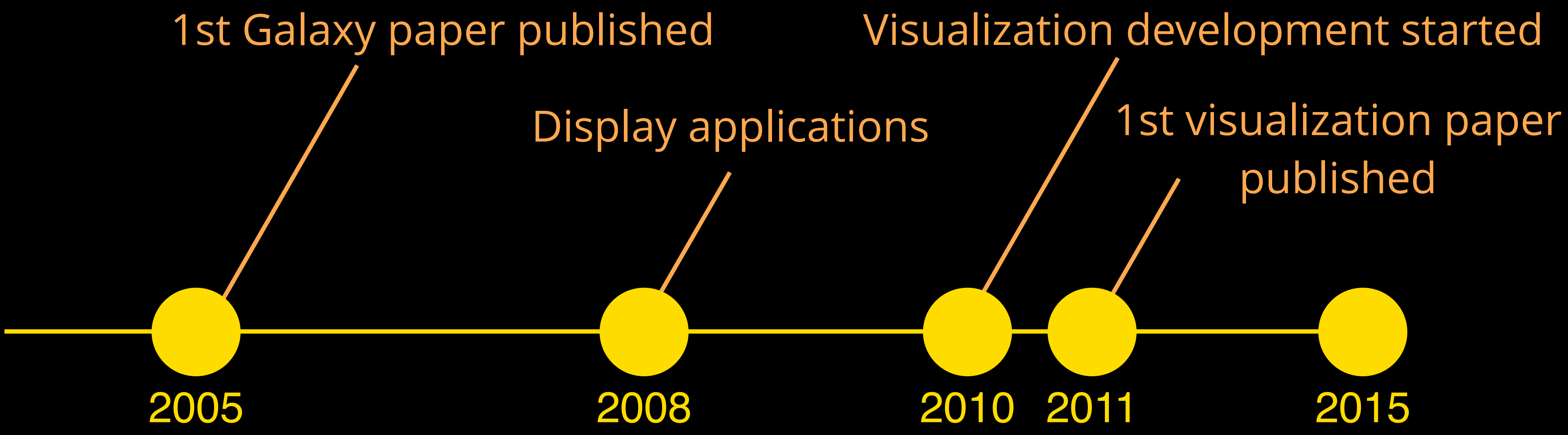
Anscombe’s Quartet



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



1st Galaxy paper published

Display applications

Visualization development started

1st visualization paper published

2005

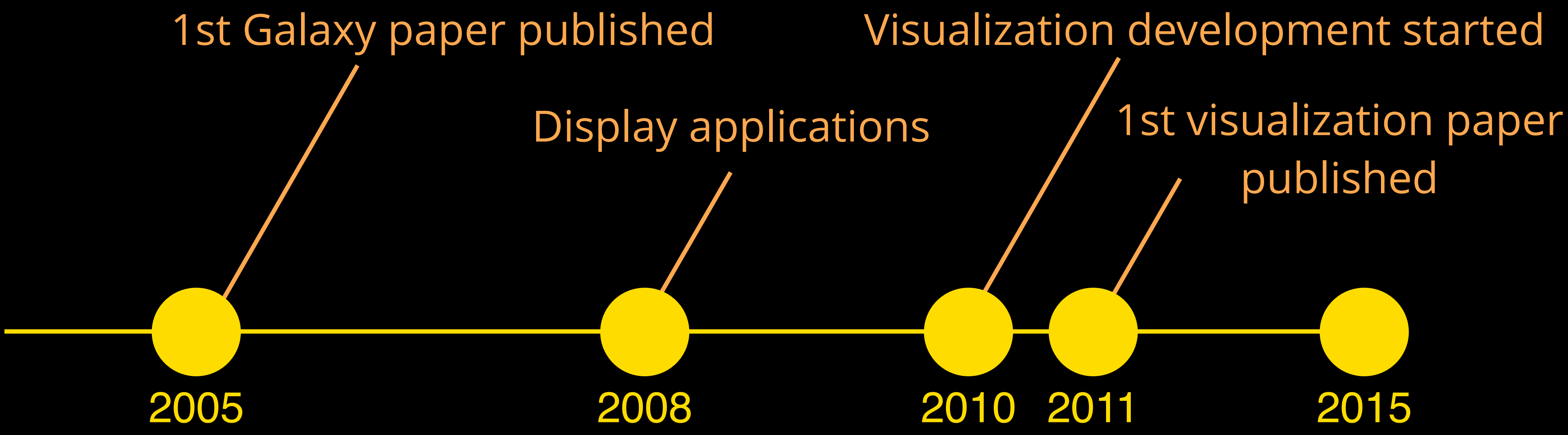
2008

2010

2011

2015

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

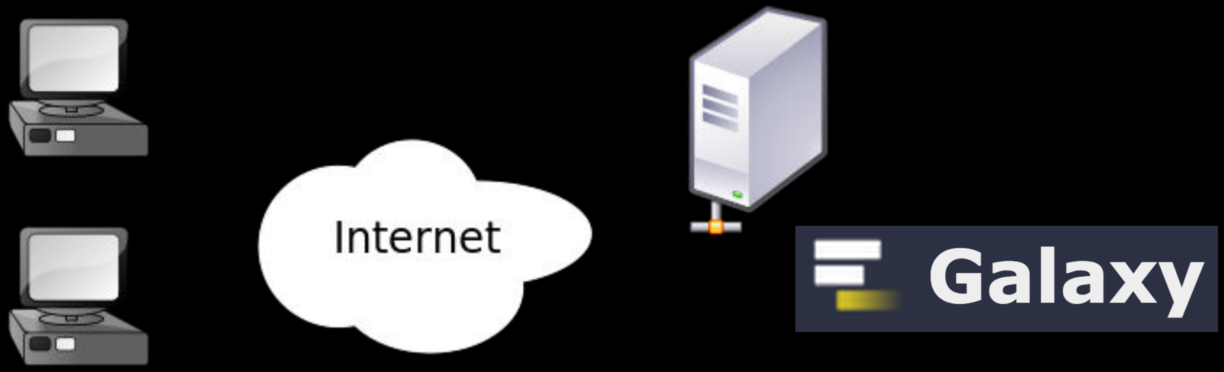
Can be saved and versioned for reproducibility

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

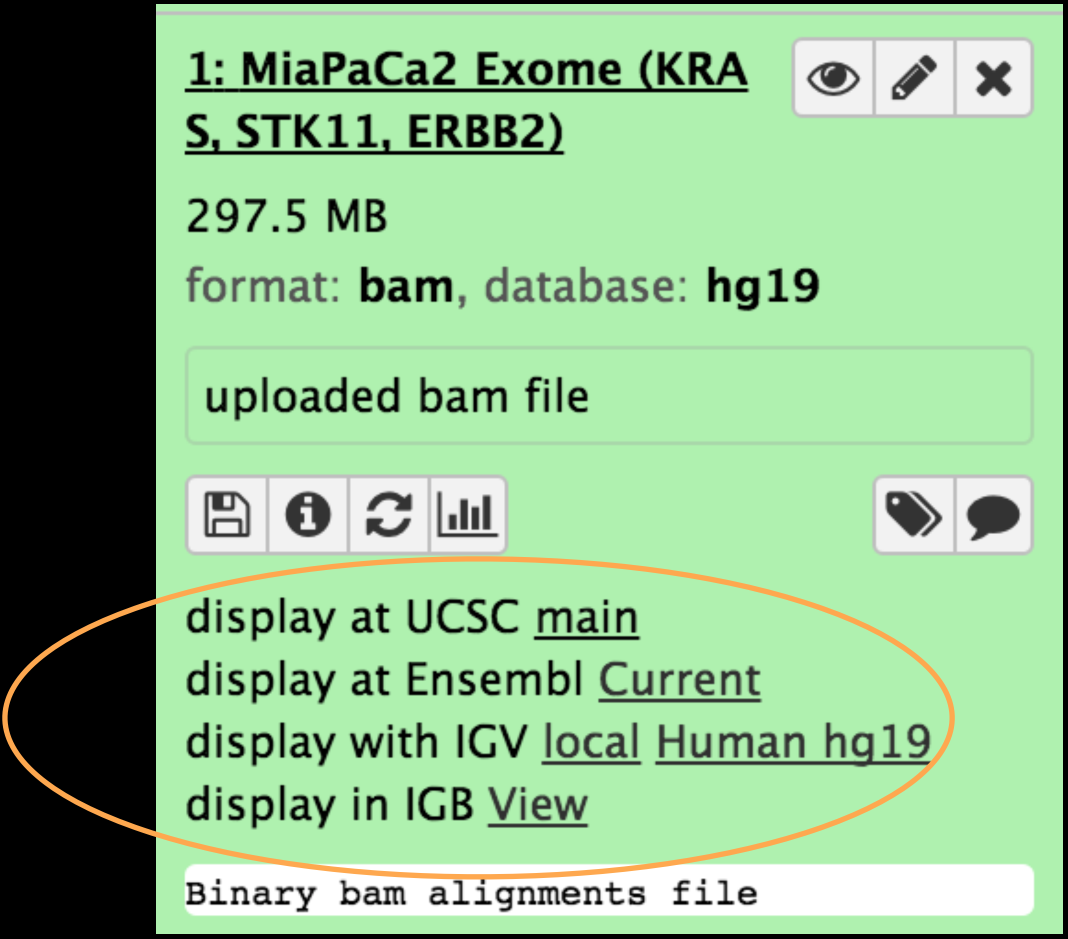
Sequenced exomes and transcriptomes of 3 pancreatic cancer cell lines

- 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 (KRAS, STK11, ERBB2)

297.5 MB

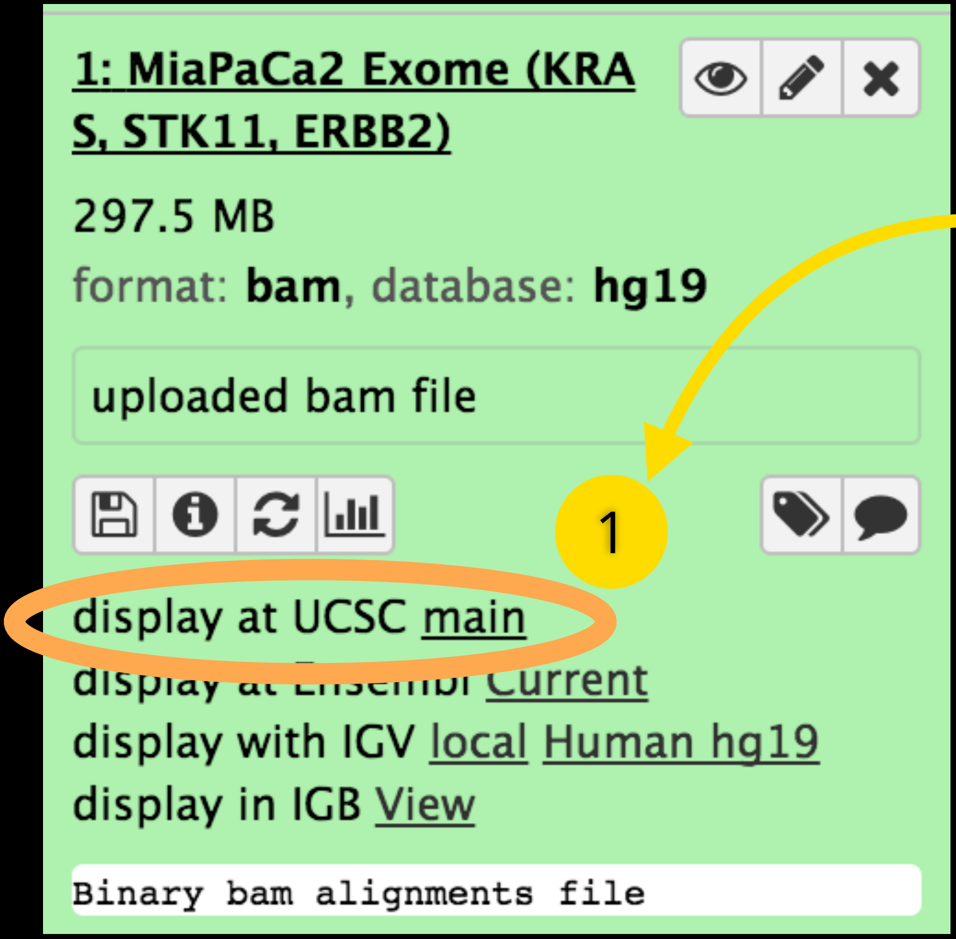
format: **bam**, database: **hg19**

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 (KRAS, STK11, ERBB2)

297.5 MB

format: **bam**, database: **hg19**

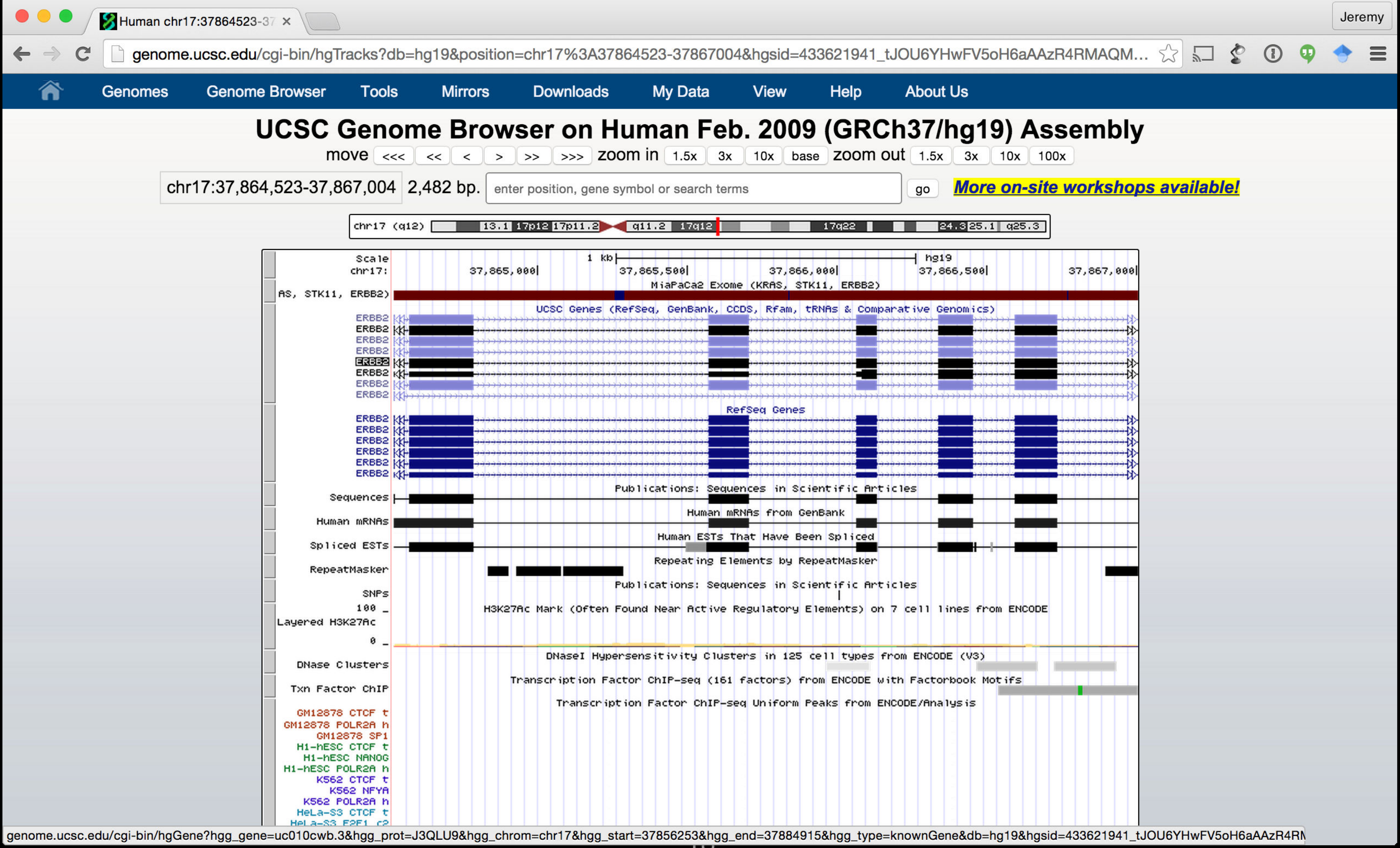
uploaded bam file

display at UCSC main
display at Ensembl Current
display with IGV local Human hg19
display in IGB View

`Binary bam alignments file`

Used throughout slides to show actions to take

Display Applications



genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr17%3A37864523-37867004&hgsid=433621941_tJOU6YHwFV5oH6aAAzR4RMAQM...

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

chr17:37,864,523-37,867,004 2,482 bp.

More on-site workshops available!

genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc010cwb.3&hgg_prot=J3QLU9&hgg_chrom=chr17&hgg_start=37856253&hgg_end=37884915&hgg_type=knownGene&db=hg19&hgsid=433621941_tJOU6YHwFV5oH6aAAzR4RM

Display Applications

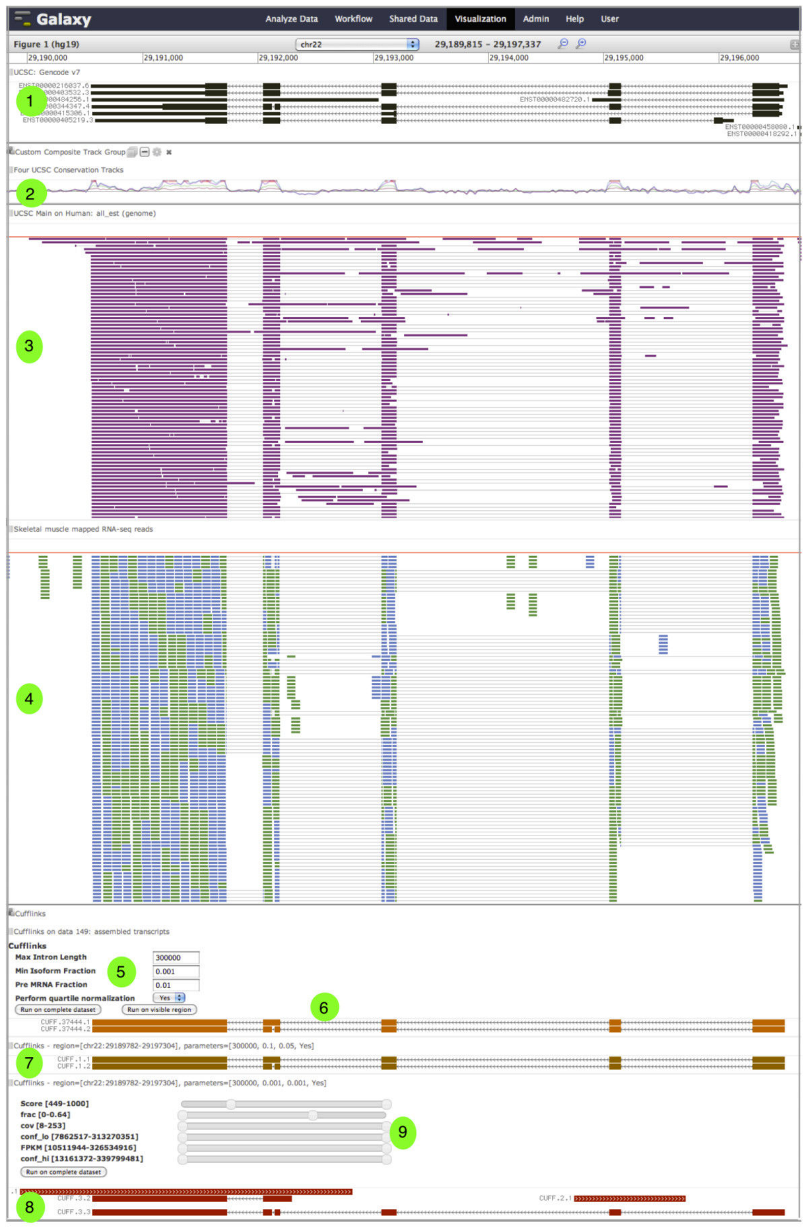
Advantages

- 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



Cufflinks

Max Intron Length 300000

Min Isoform Fraction 0.001

Pre MRNA Fraction 0.01

Perform quartile normalization Yes

Score [449-1000]

frac [0-0.64]

cov [8-253]

conf_lo [7862517-313270351]

FPKM [10511944-326534916]

conf_hi [13161372-339799481]

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

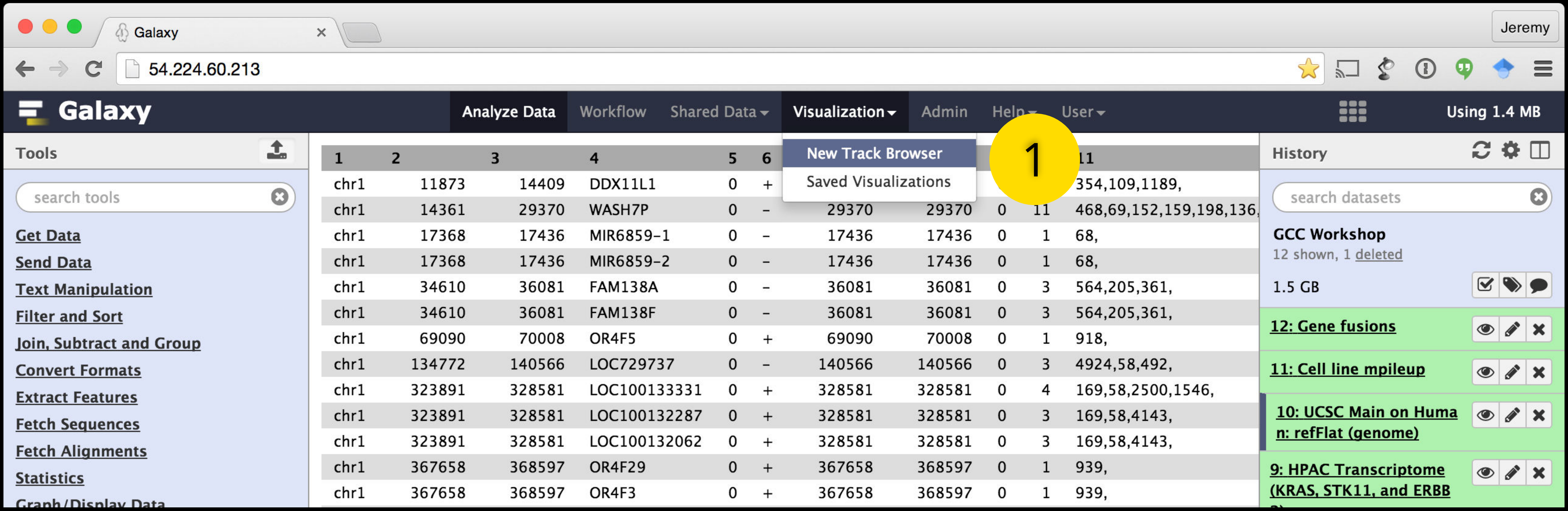
BED, GFF/GTF, interval, SAM/BAM, VCF, Wiggle, BigWig, BigBed, BedGraph

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

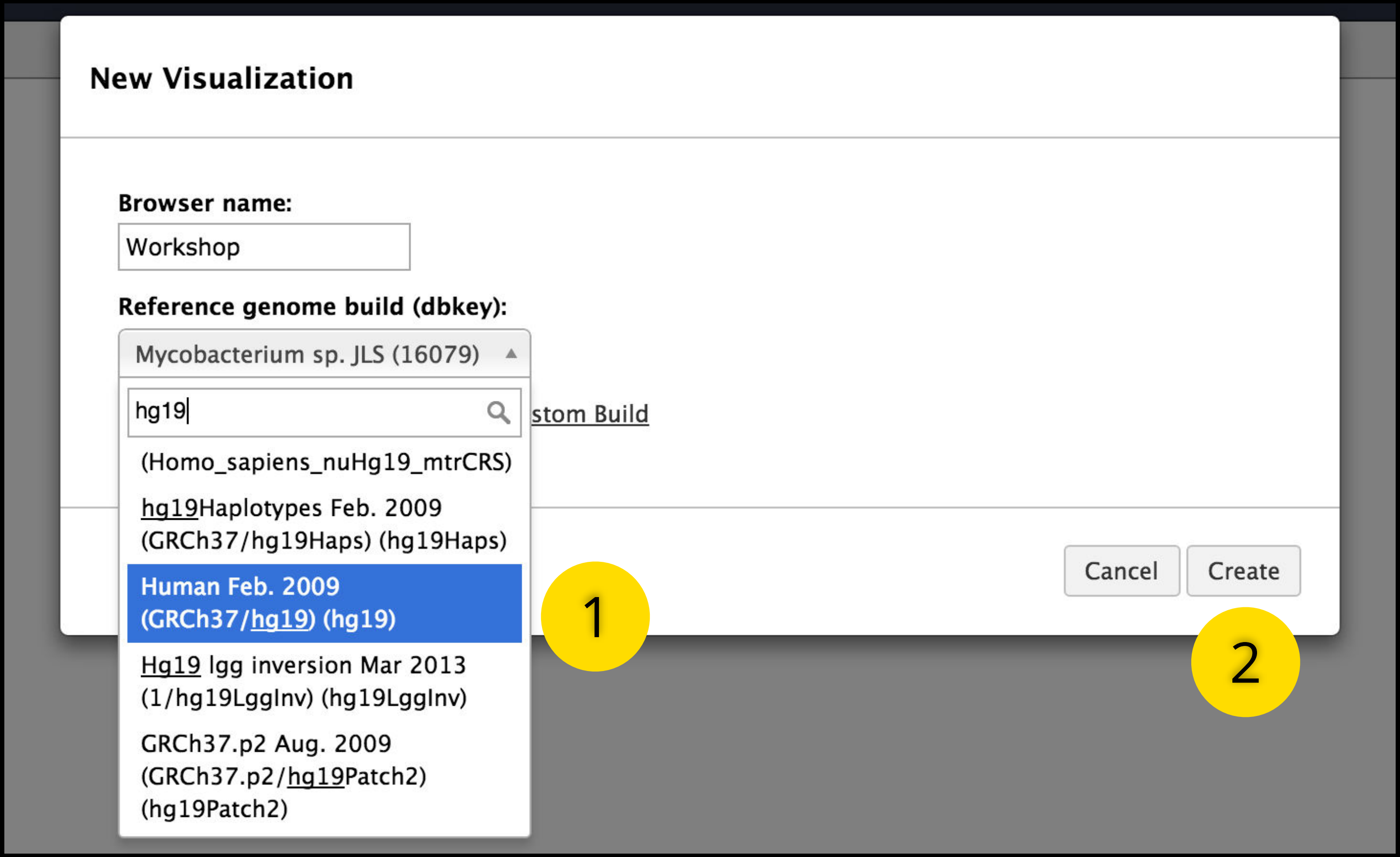
1. Create visualization



1	2	3	4	5	6					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



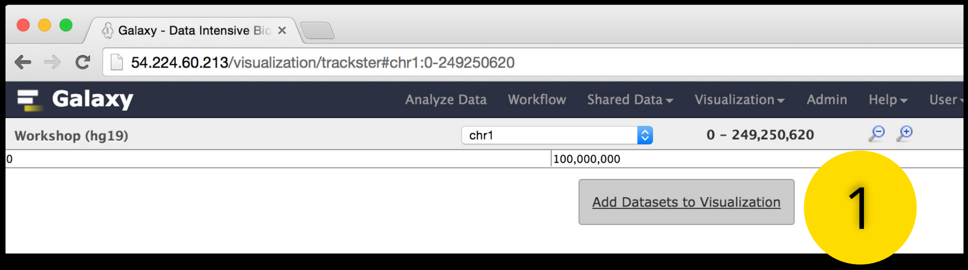
Browser name: Workshop

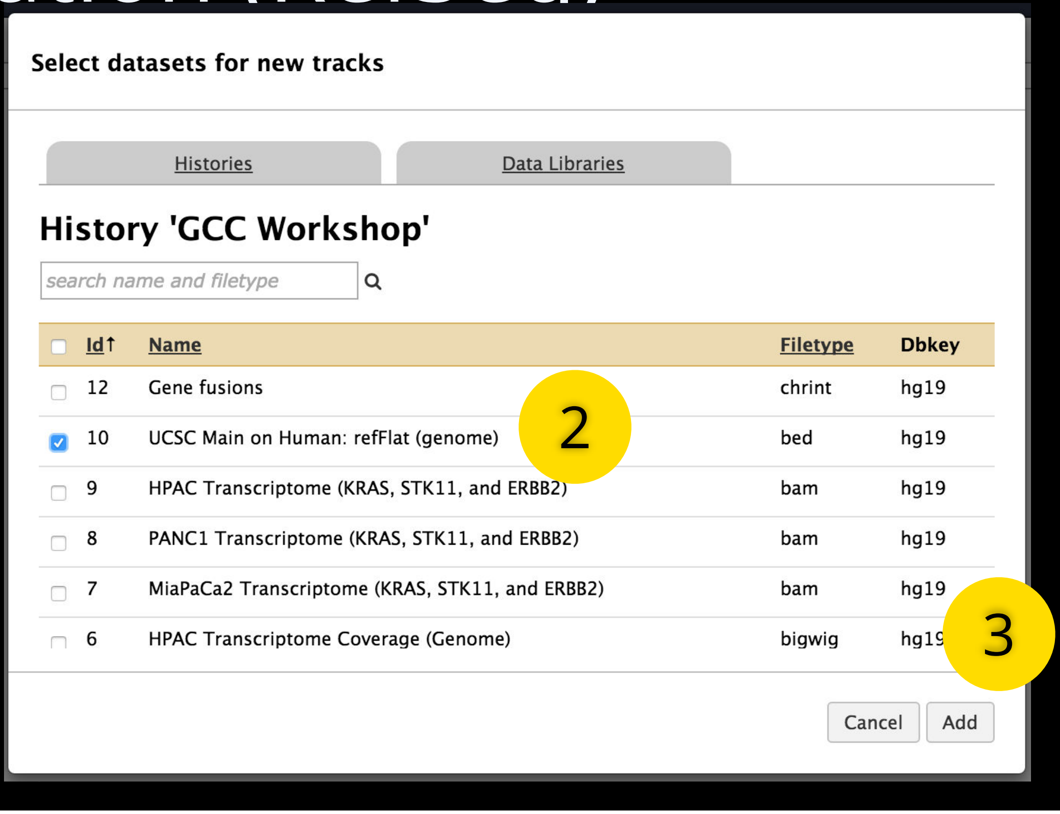
Reference genome build (dbkey): Mycobacterium sp. JLS (16079)

hg19

Let’s visualize our data in Trackster

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





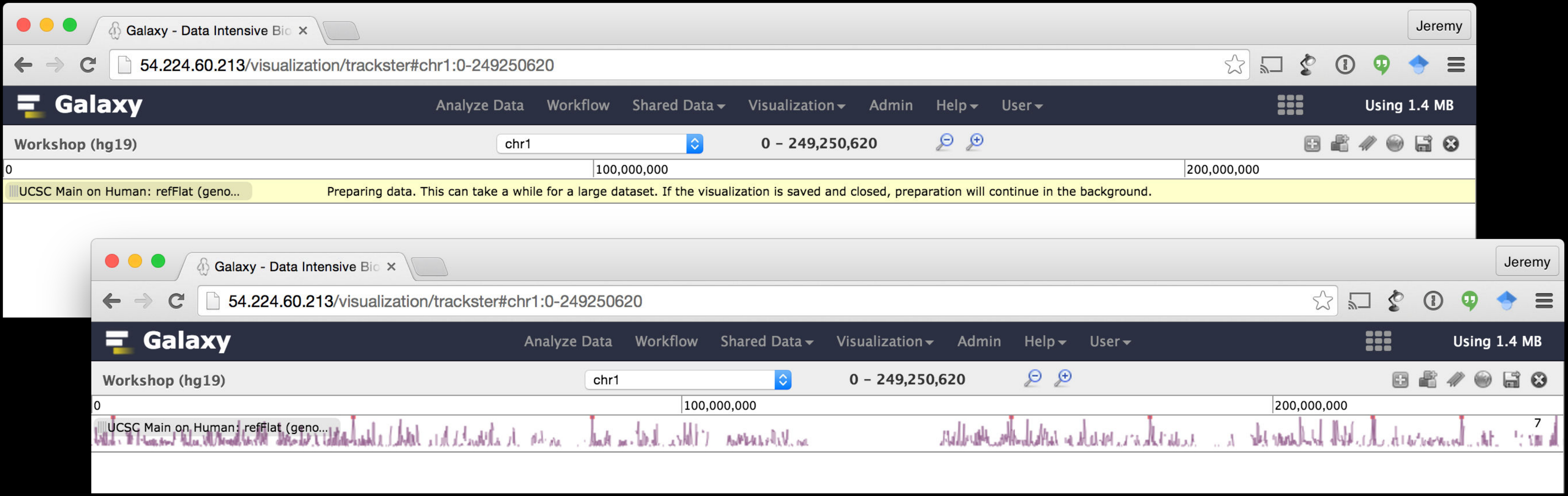
Select datasets for new tracks

History 'GCC Workshop'

	Id	Name	Filetype	Dbkey
☐	12	Gene fusions	chrint	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

Let’s visualize our data in Trackster

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



54.224.60.213/visualization/trackster#chr1:0-249250620

Workshop (hg19)

chr1

0 – 249,250,620

UCSC Main on Human: refFlat (geno...

Preparing data. This can take a while for a large dataset. If the visualization is saved and closed, preparation will continue in the background.

54.224.60.213/visualization/trackster#chr1:0-249250620

Workshop (hg19)

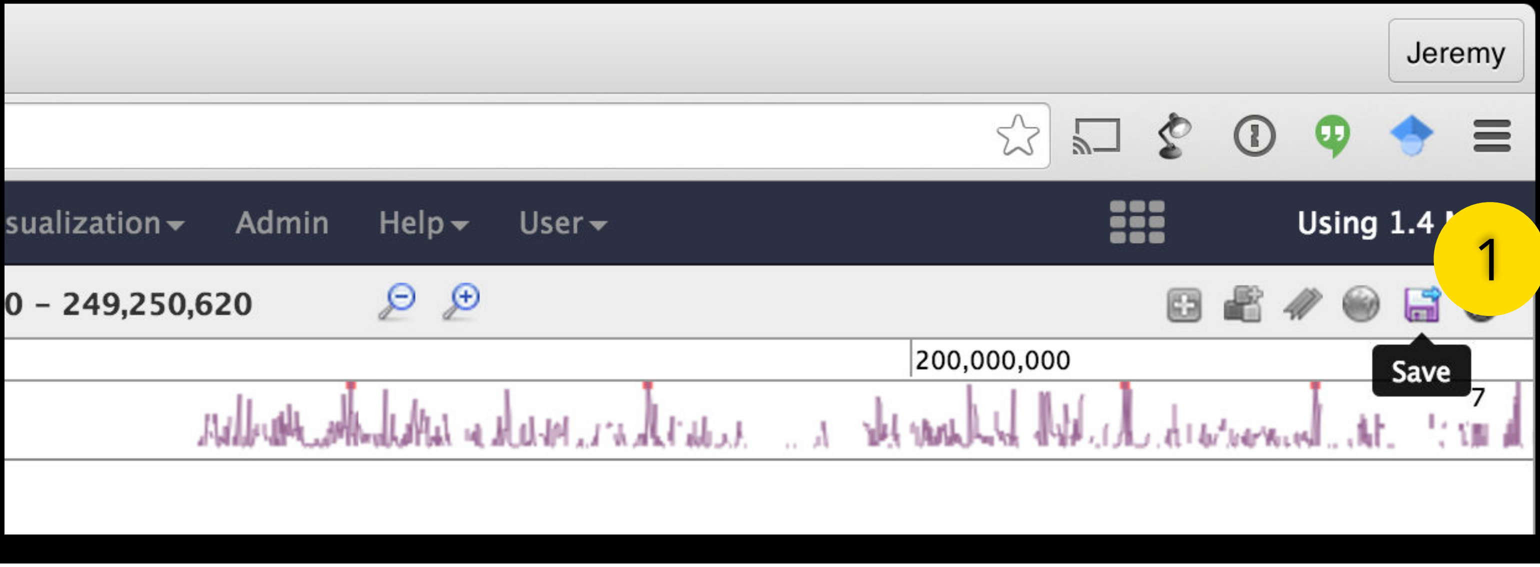
chr1

0 – 249,250,620

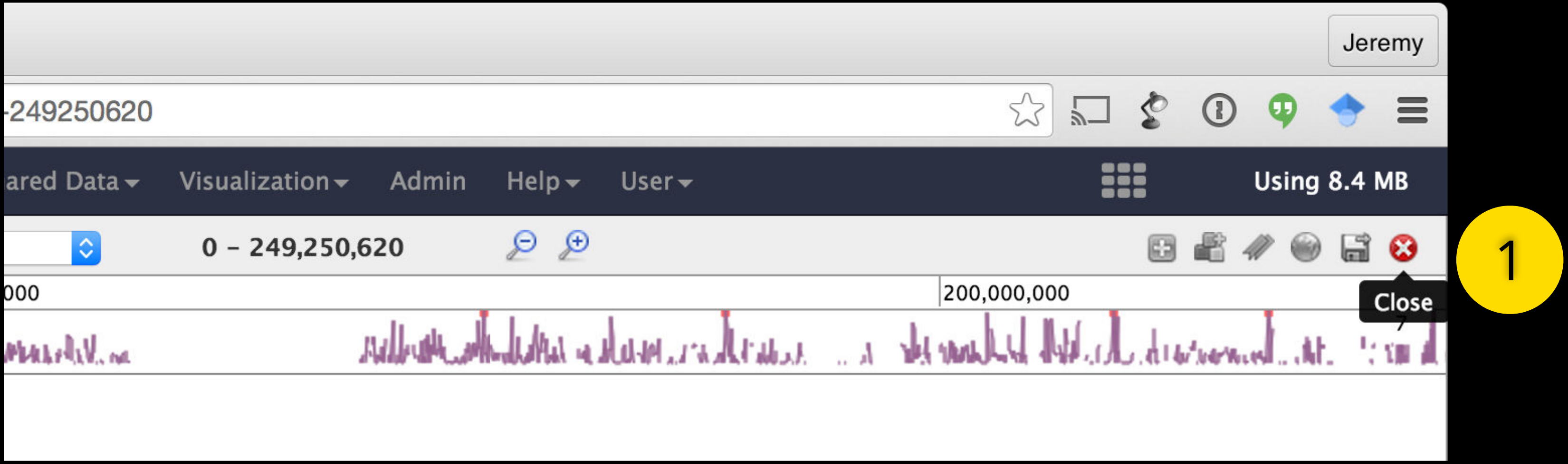
UCSC Main on Human: refFlat (geno...

Let's visualize our data in Trackster

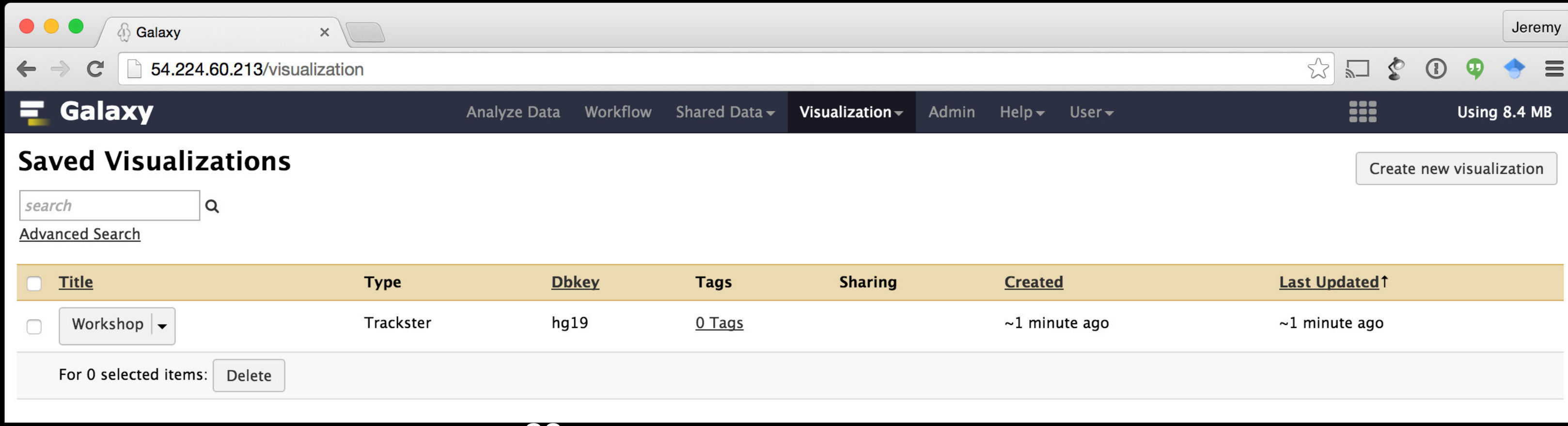
1. Create visualization
2. Add gene annotation (RefSeq)
3. Save visualization



Let's visualize our data in Trackster



4. Exit

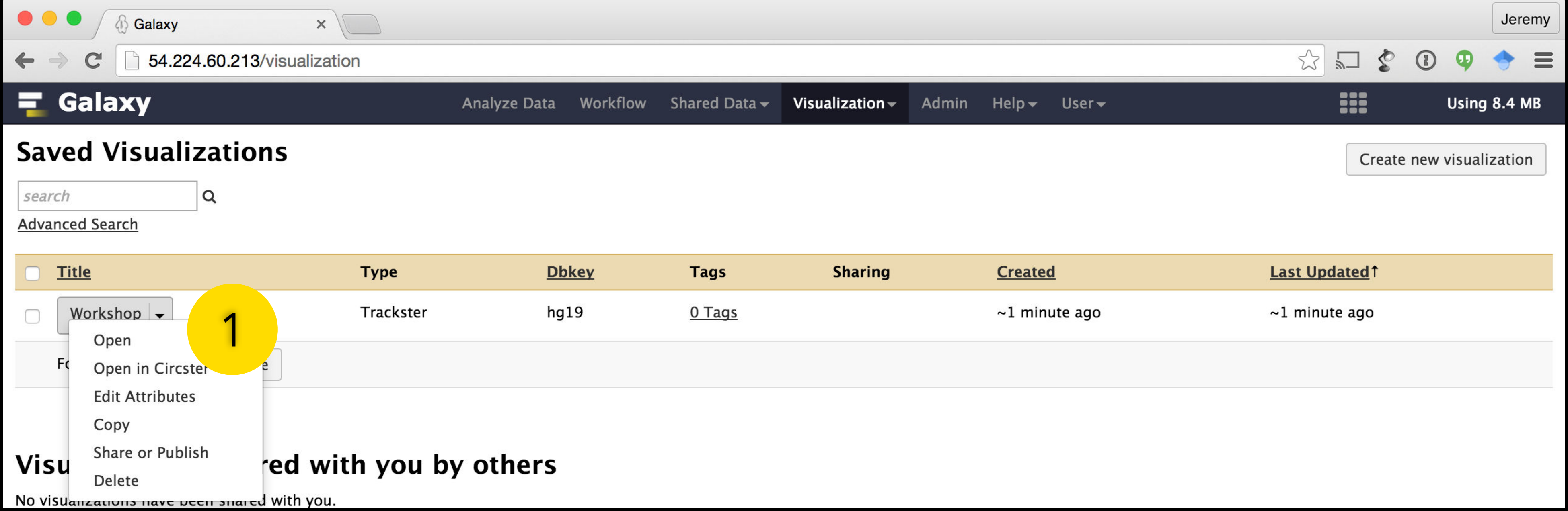


Saved Visualizations

☐	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



Saved Visualizations

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

Visu… …red with you by others

No visu… …with you.

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

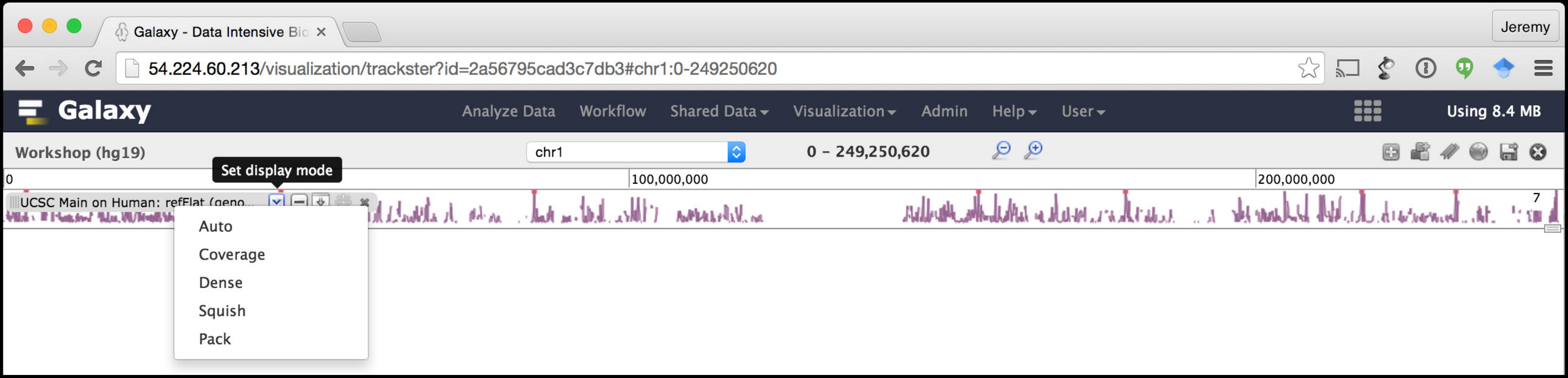
Display Modes

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

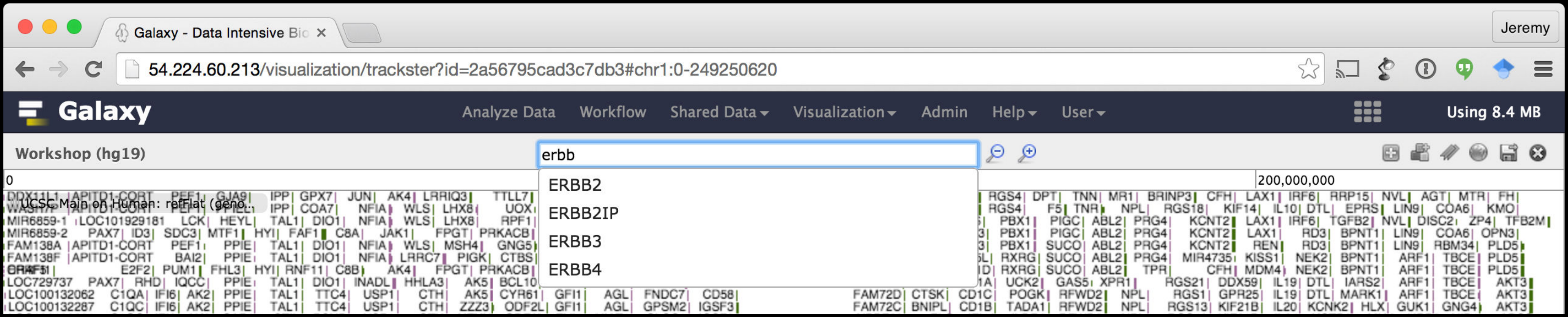


Searching

Can search for named features such as gene annotations

- BED, GFF/GTF

Let’s try searching for a gene: ERBB2

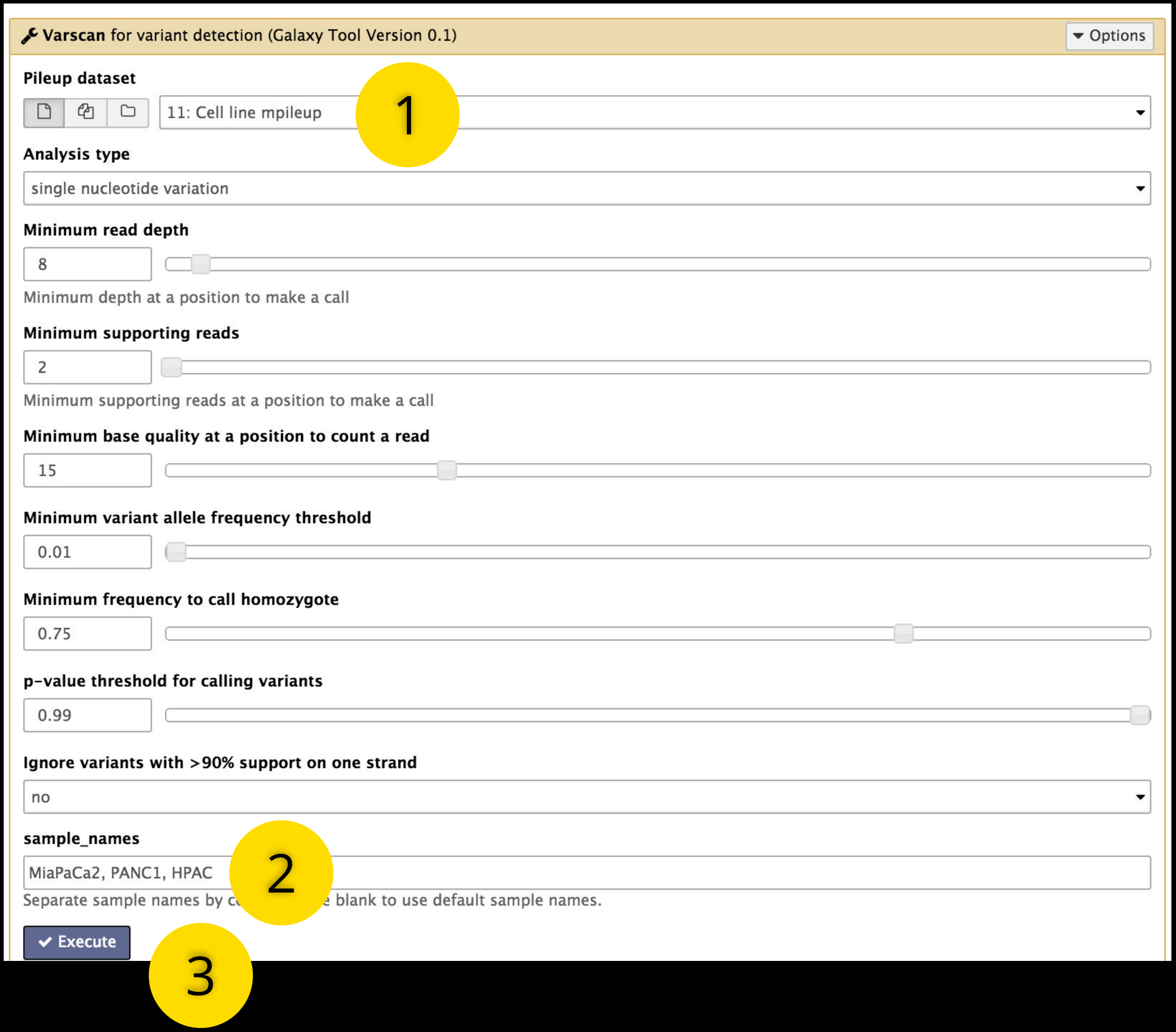


Let’s Call Variants

VarScan

- Sample names: MiaPaCa2, PANC1, HPAC
- Run

Rename output: “Cell line variants”



Varscan for variant detection (Galaxy Tool Version 0.1)

Pileup dataset
11: Cell line mpileup

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

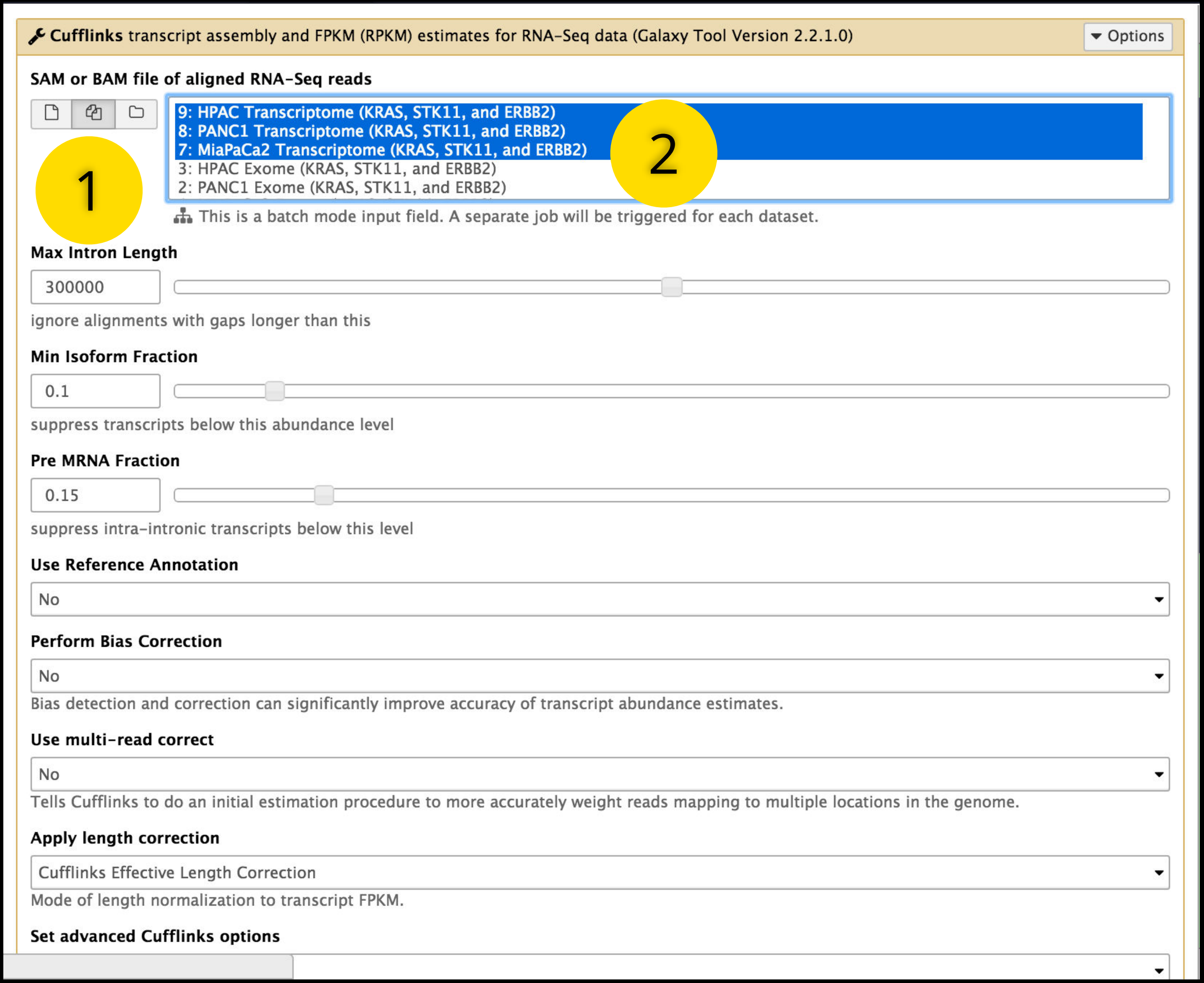
sample_names
MiaPaCa2, PANC1, HPAC
Separate sample names by c e blank to use default sample names.

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)

SAM or BAM file of aligned RNA-Seq reads

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

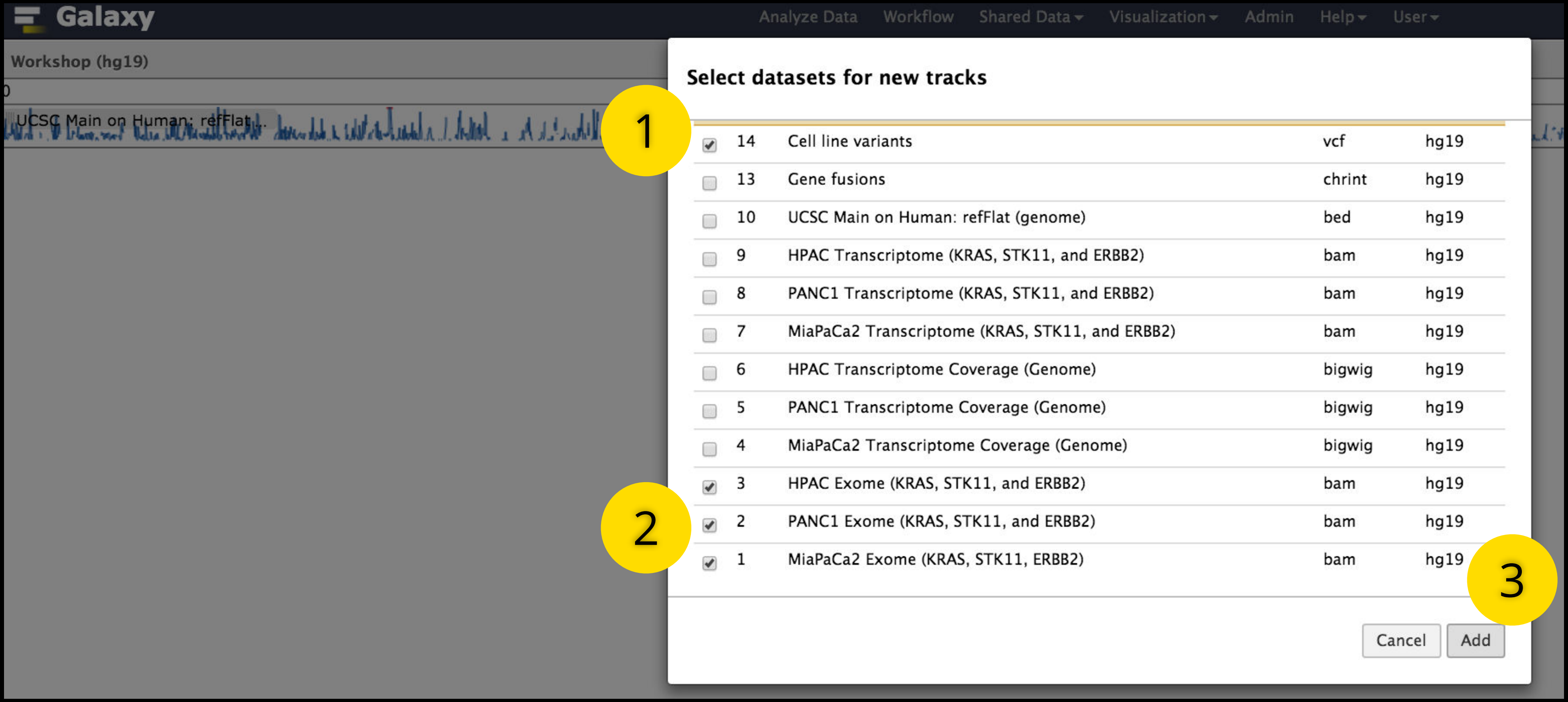
Cufflinks Effective Length Correction

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…



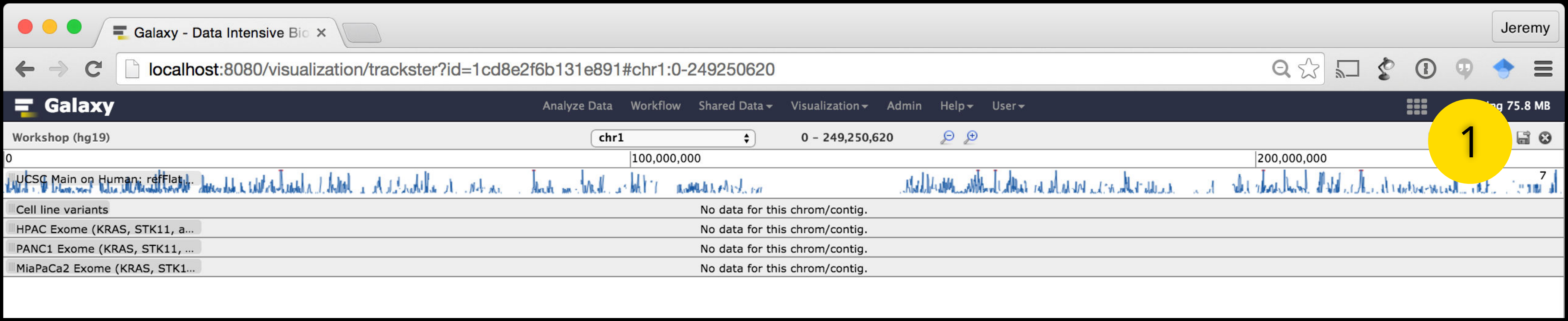
Select datasets for new tracks

☑	14	Cell line variants	vcf	hg19
☐	13	Gene fusions	chrint	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

Let’s add data to Trackster

Add exome data for all cell lines and called variants…

…but where is our data?



Workshop (hg19)

chr1

0 - 249,250,620

UCSC Main on Human: refFlat...

Cell line variants: No data for this chrom/contig.

HPAC Exome (KRAS, STK11, a...: No data for this chrom/contig.

PANC1 Exome (KRAS, STK11, ...: No data for this chrom/contig.

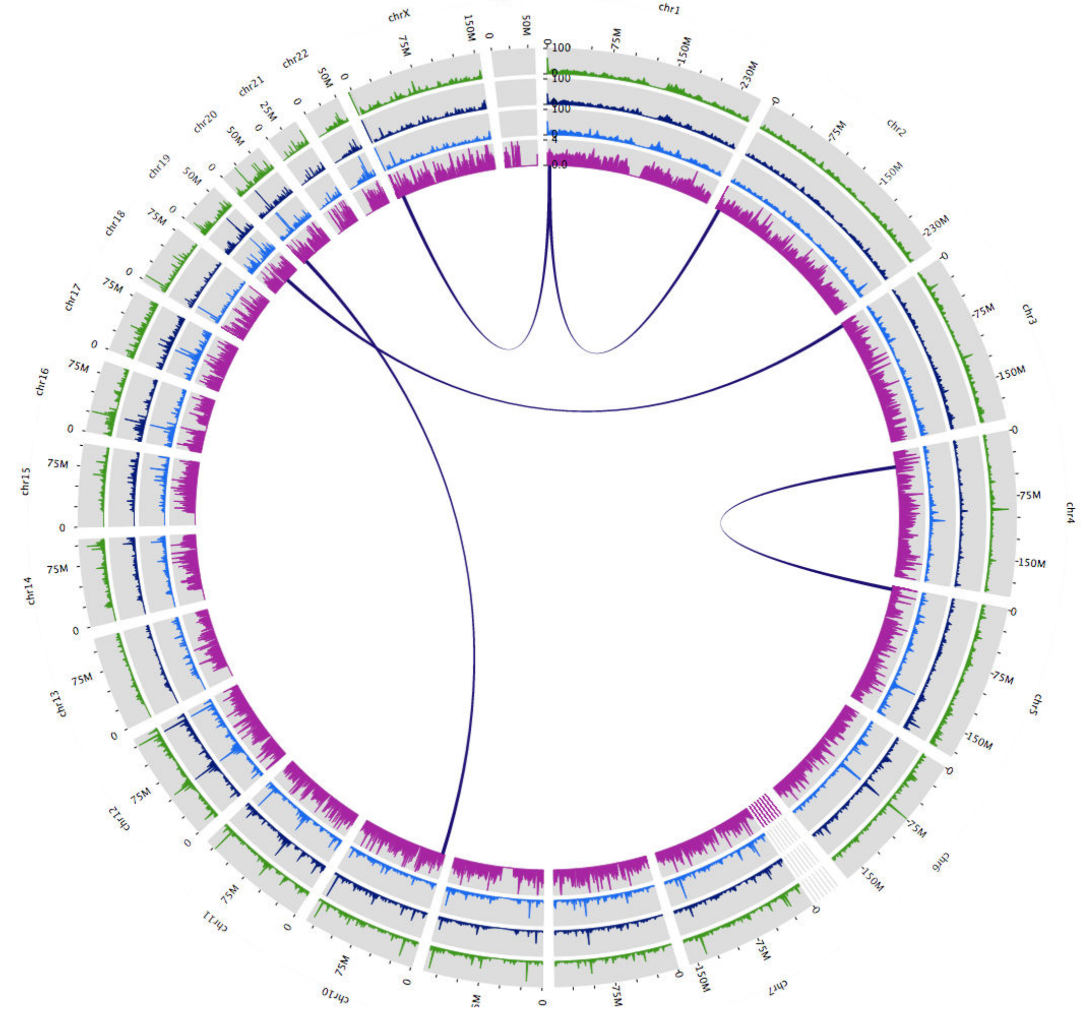
MiaPaCa2 Exome (KRAS, STK1...: No data for this chrom/contig.

Save again

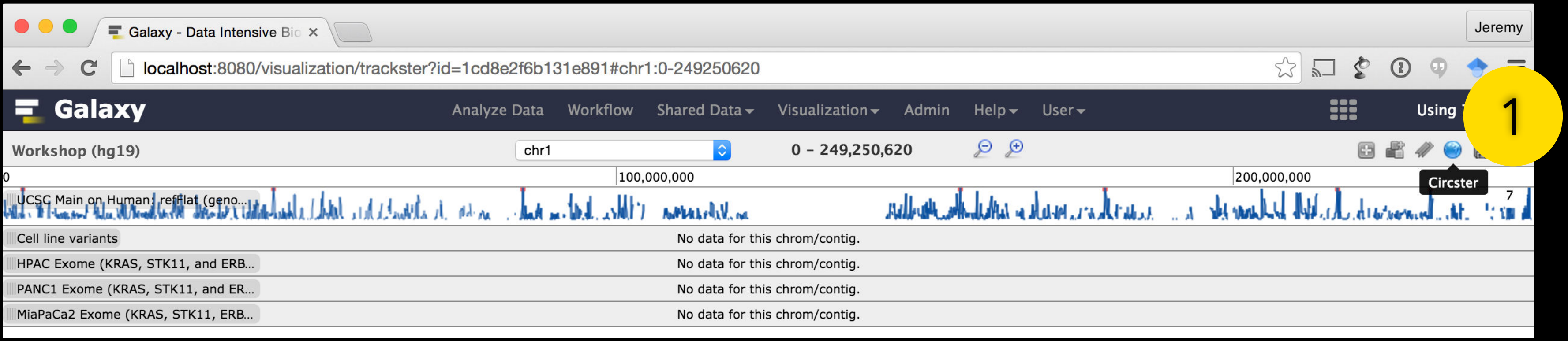
Circster

Interactive Circos plot

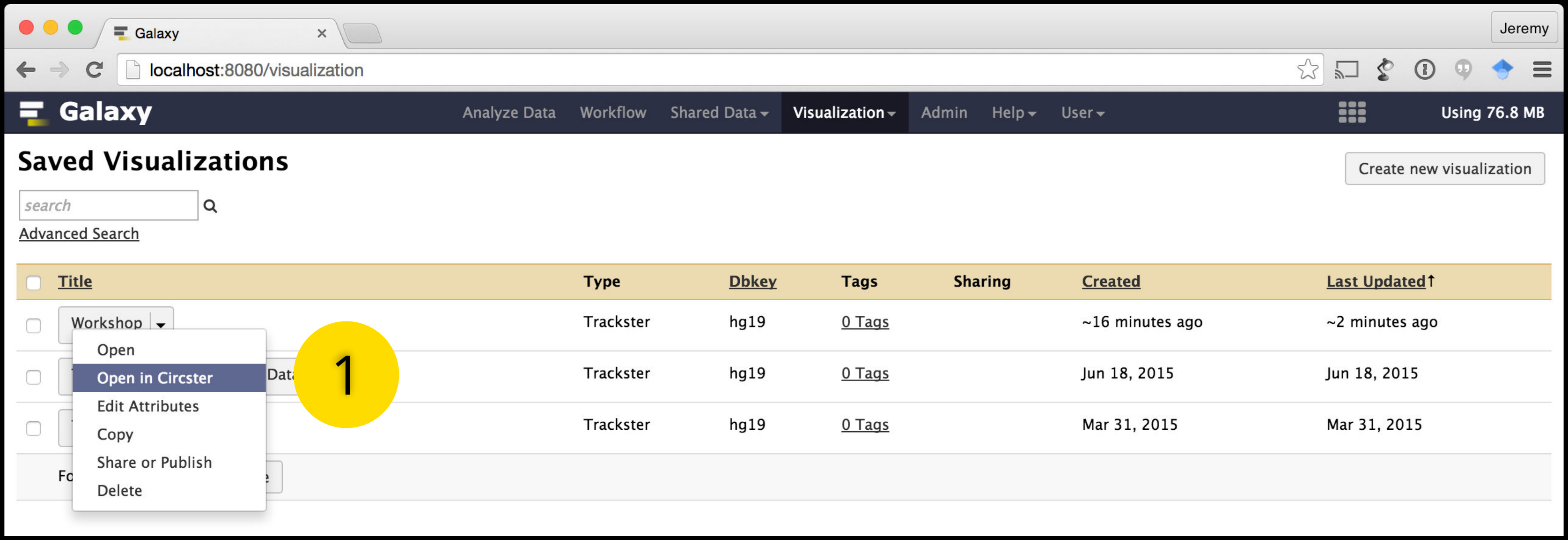
Whole genome view with structural variation



Let’s view our data in Circster



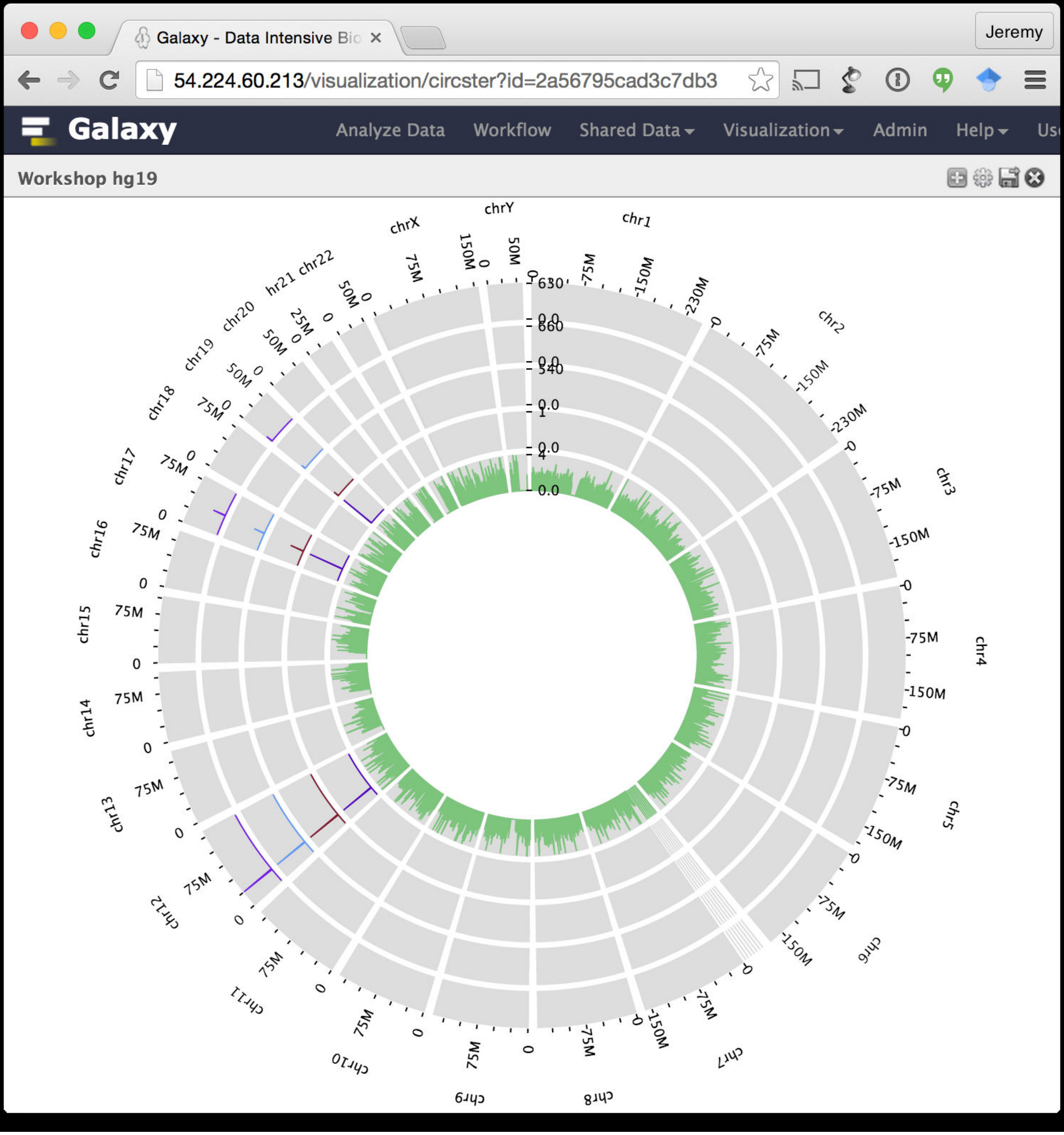
UCSC Main on Human: refFlat (geno...	
Cell line variants	No data for this chrom/contig.
HPAC Exome (KRAS, STK11, and ERB...	No data for this chrom/contig.
PANC1 Exome (KRAS, STK11, and ER...	No data for this chrom/contig.
MiaPaCa2 Exome (KRAS, STK11, ERB...	No data for this chrom/contig.



Saved Visualizations

☐	Title	Type	Dbkey	Tags	Sharing	Created	Last Updated↑
☐	Workshop	Trackster	hg19	0 Tags		~16 minutes ago	~2 minutes ago
☐	Dat	Trackster	hg19	0 Tags		Jun 18, 2015	Jun 18, 2015
☐		Trackster	hg19	0 Tags		Mar 31, 2015	Mar 31, 2015

Let’s view our data in Circster

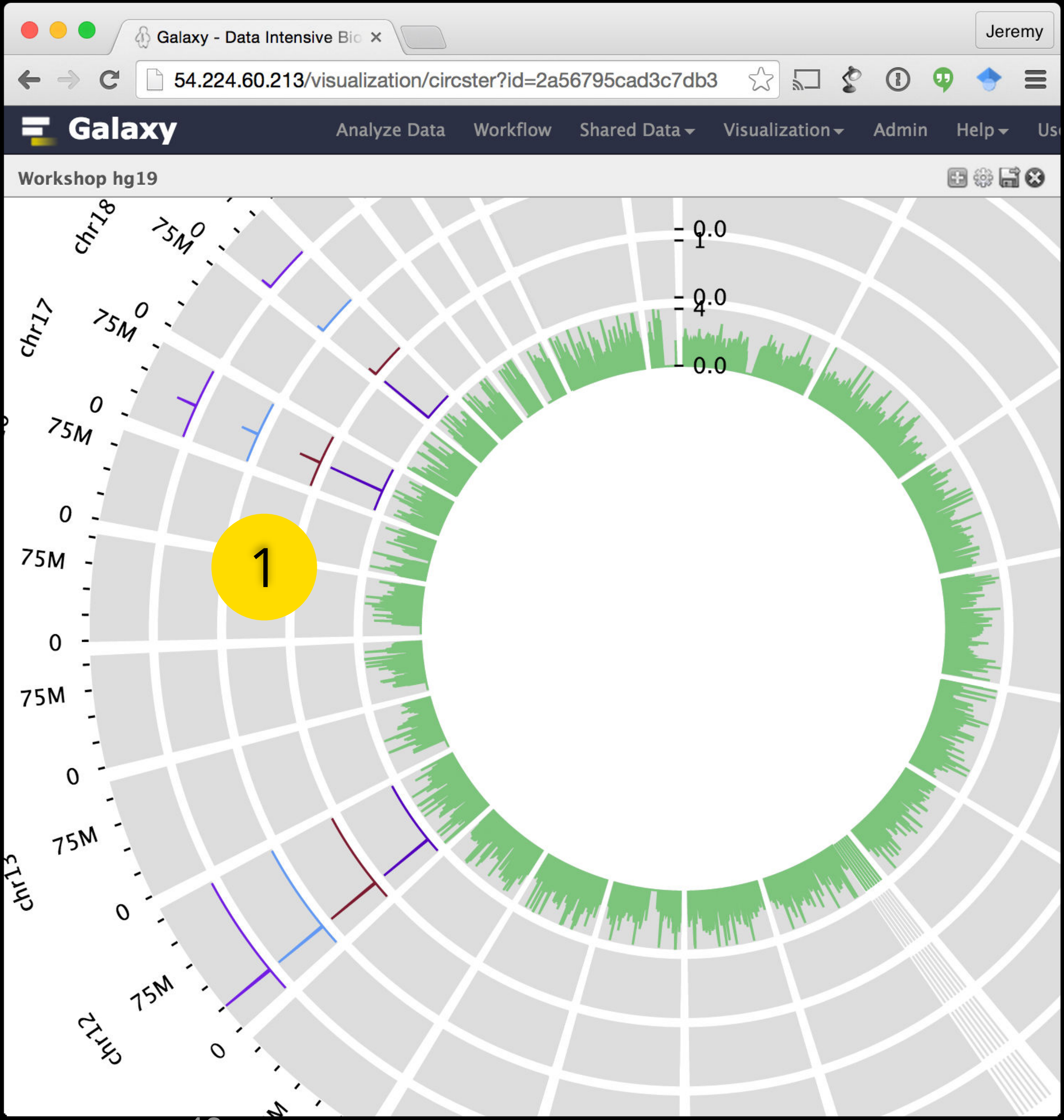


Let's view our data in Circster

Double-click or use trackpad to zoom in

Drag around using mouse/trackpad

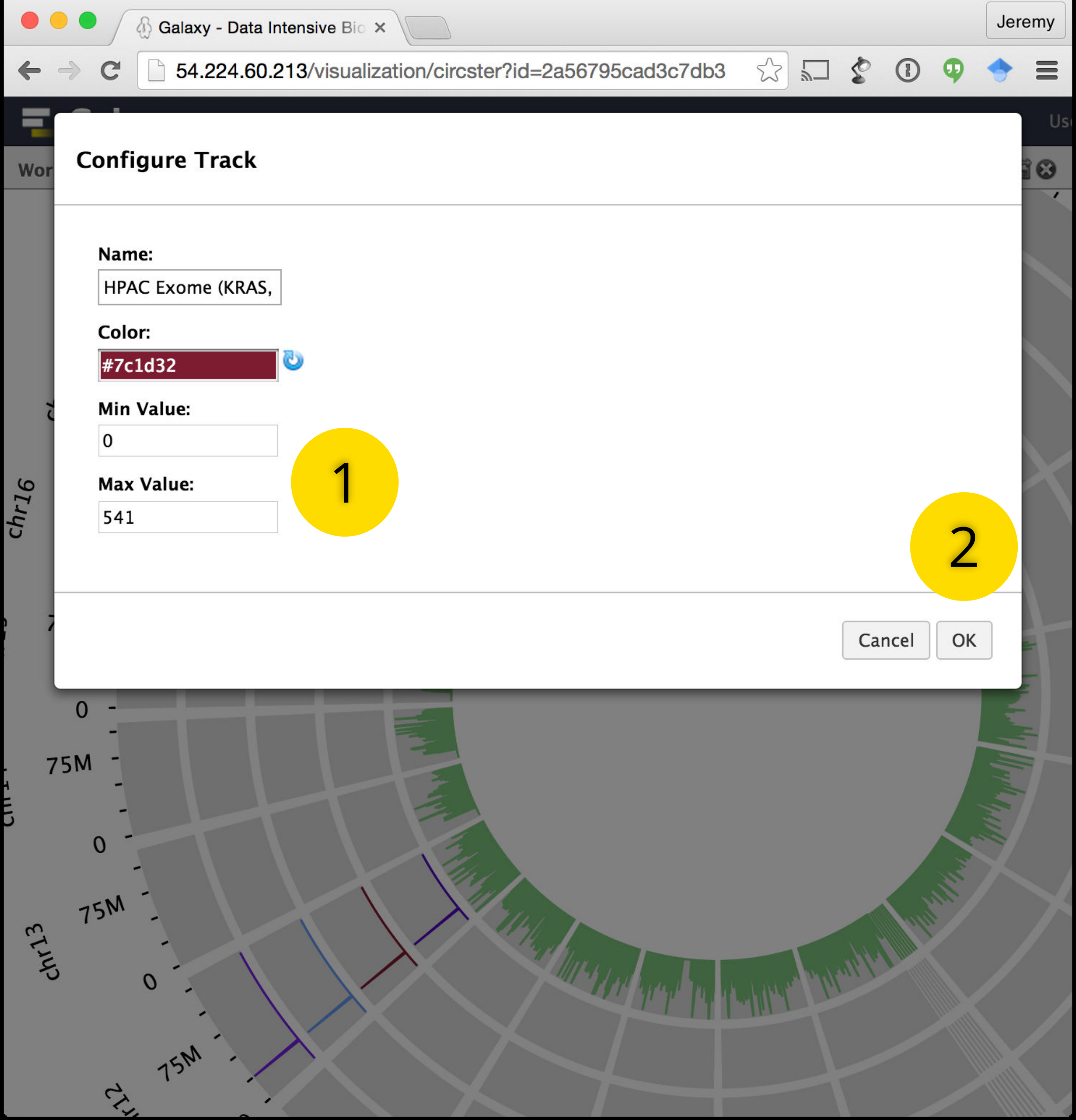
What do we see?



Let’s view our data in Circster

Change min/max by clicking on labels

What do we see?



Name: HPAC Exome (KRAS,

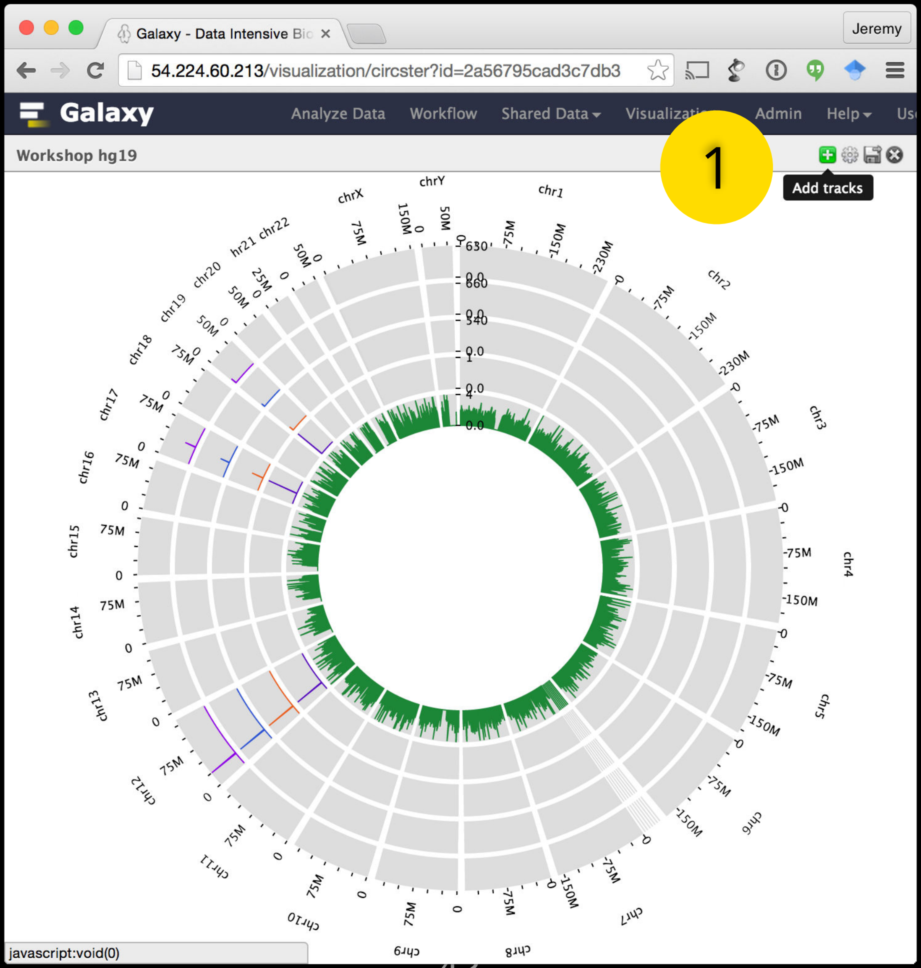
Color: #7c1d32

Min Value: 0

Max Value: 541

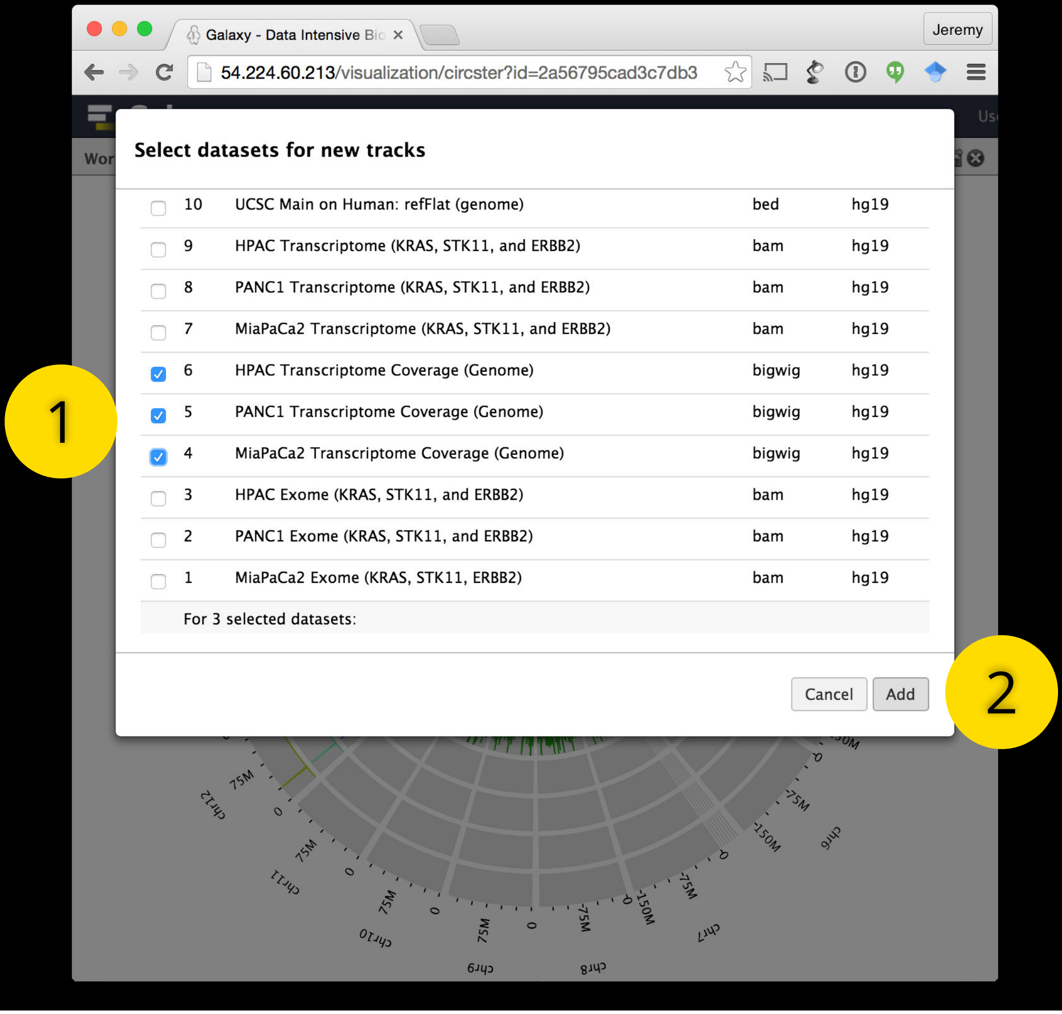
Let's add data to Circster and adjust options

1. Add transcriptome coverage data



Let’s add data to Circster and adjust options

1. Add transcriptome coverage data



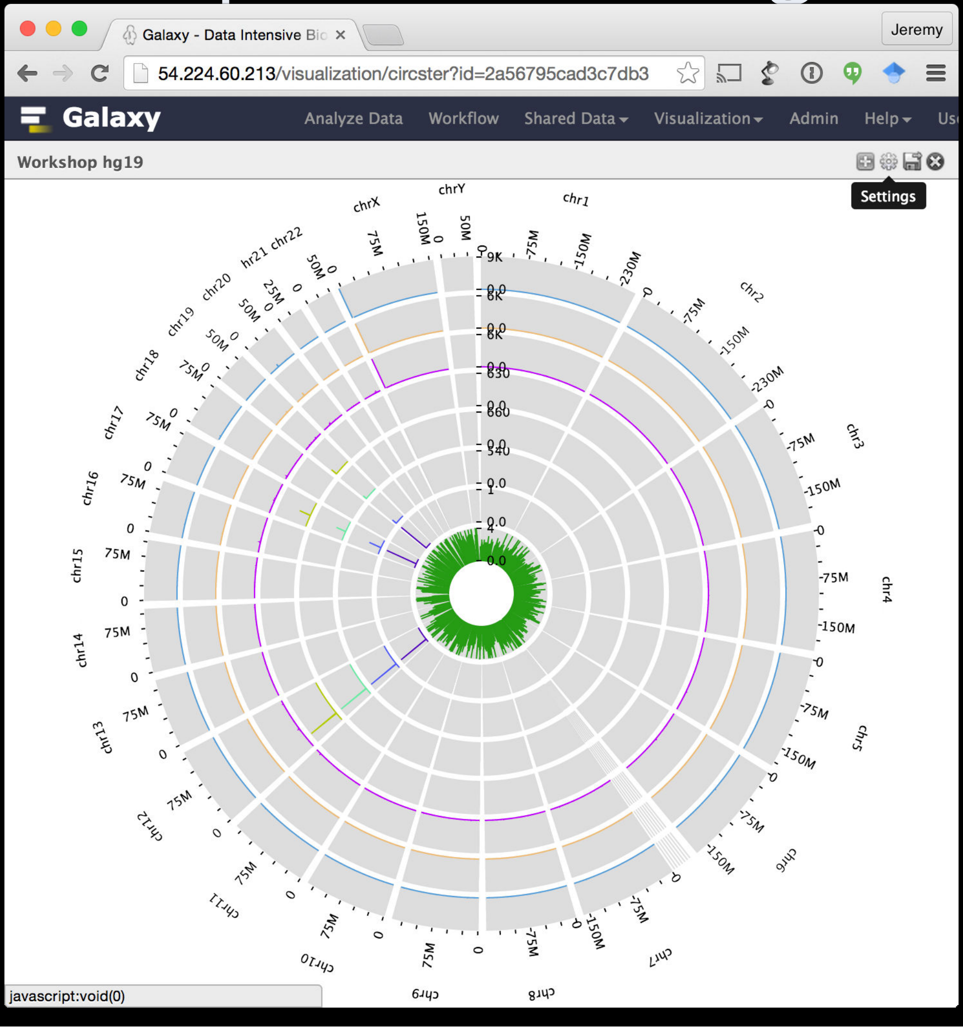
Select datasets for new tracks

☐	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

For 3 selected datasets:

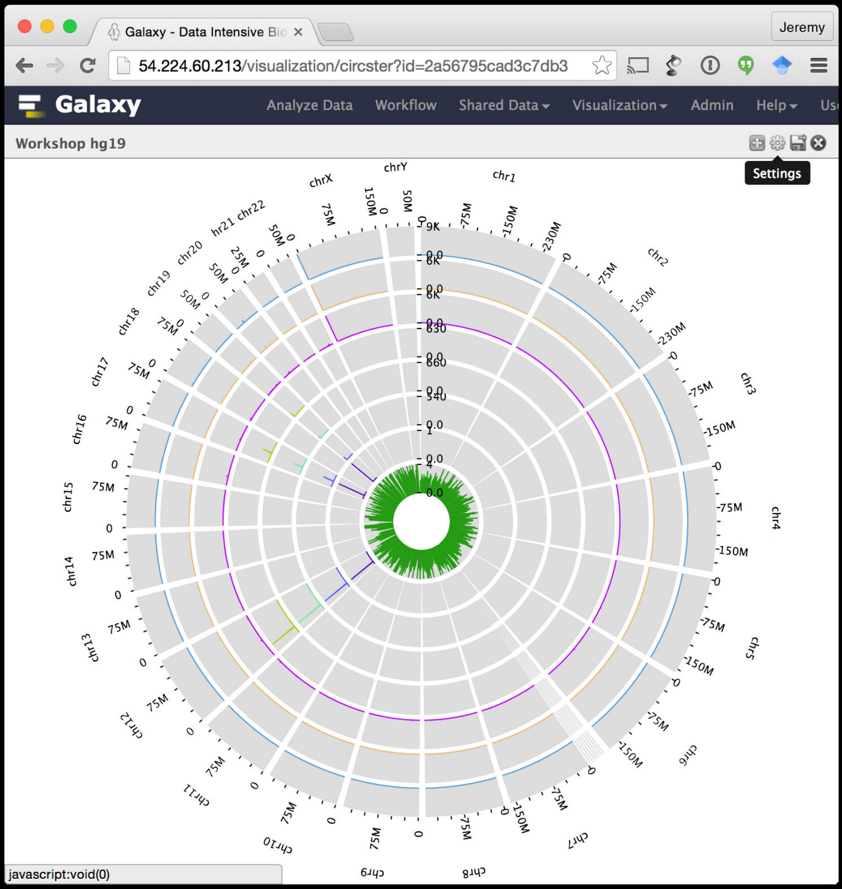
Let’s add data to Circster and adjust options

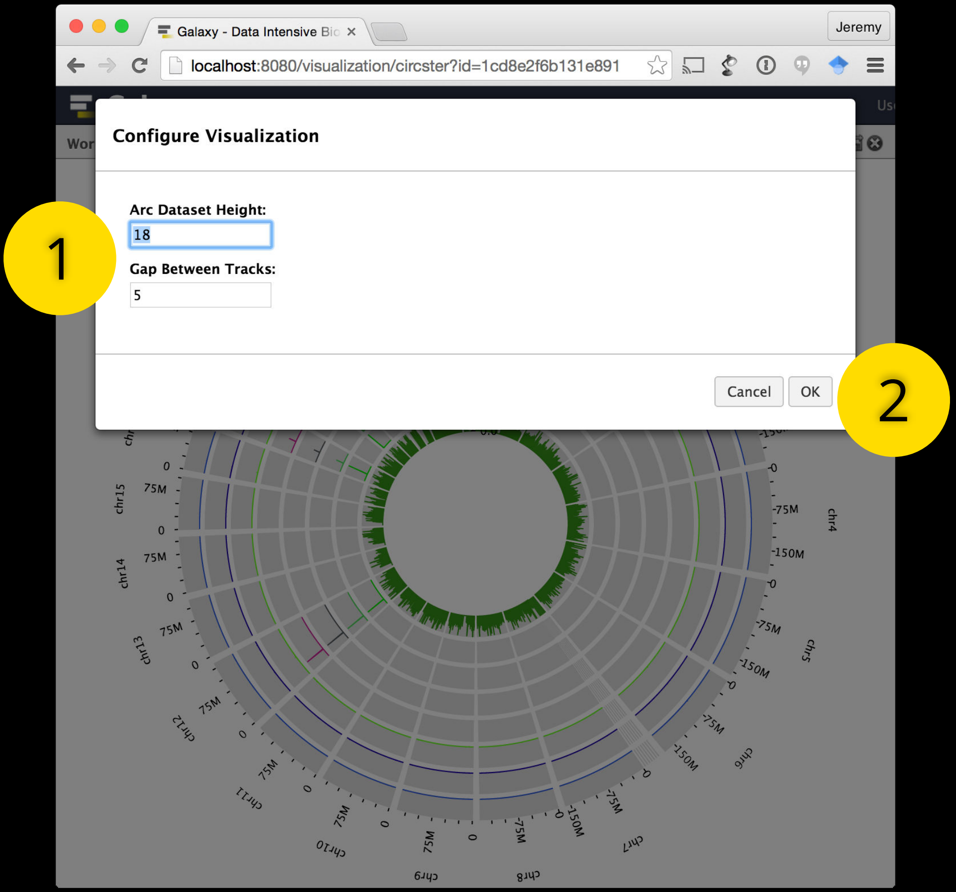
1. Add transcriptome coverage data



Let’s add data to Circster and adjust options

2. Change arc dataset height





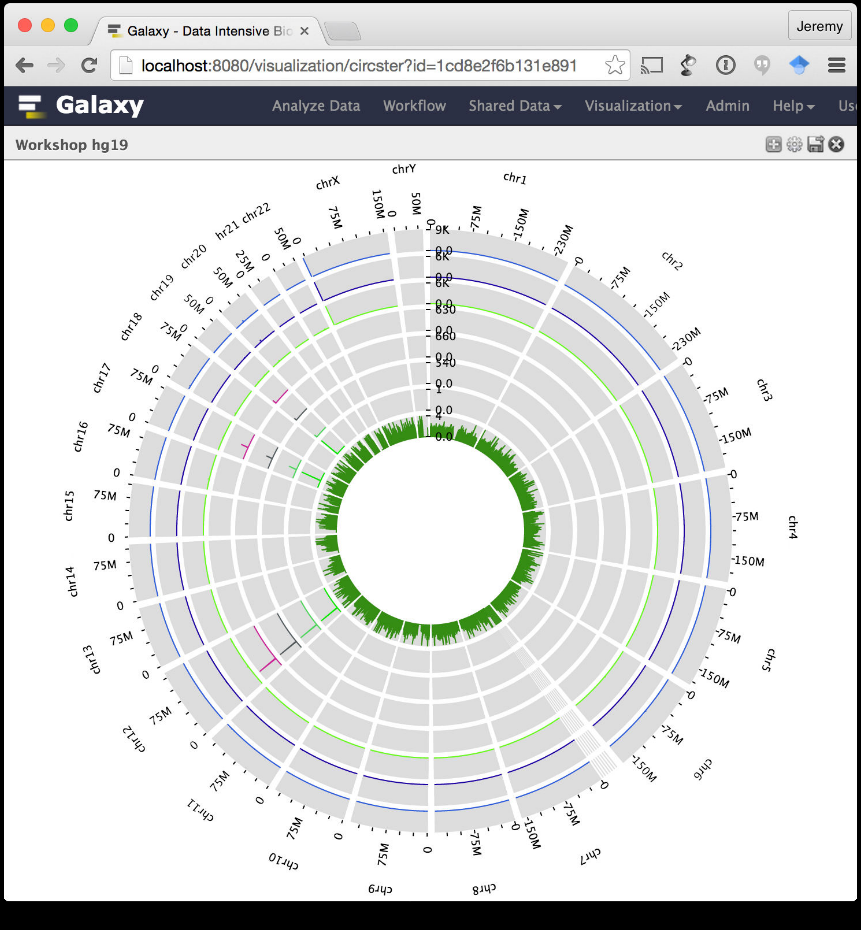
Configure Visualization

Arc Dataset Height: 18

Gap Between Tracks: 5

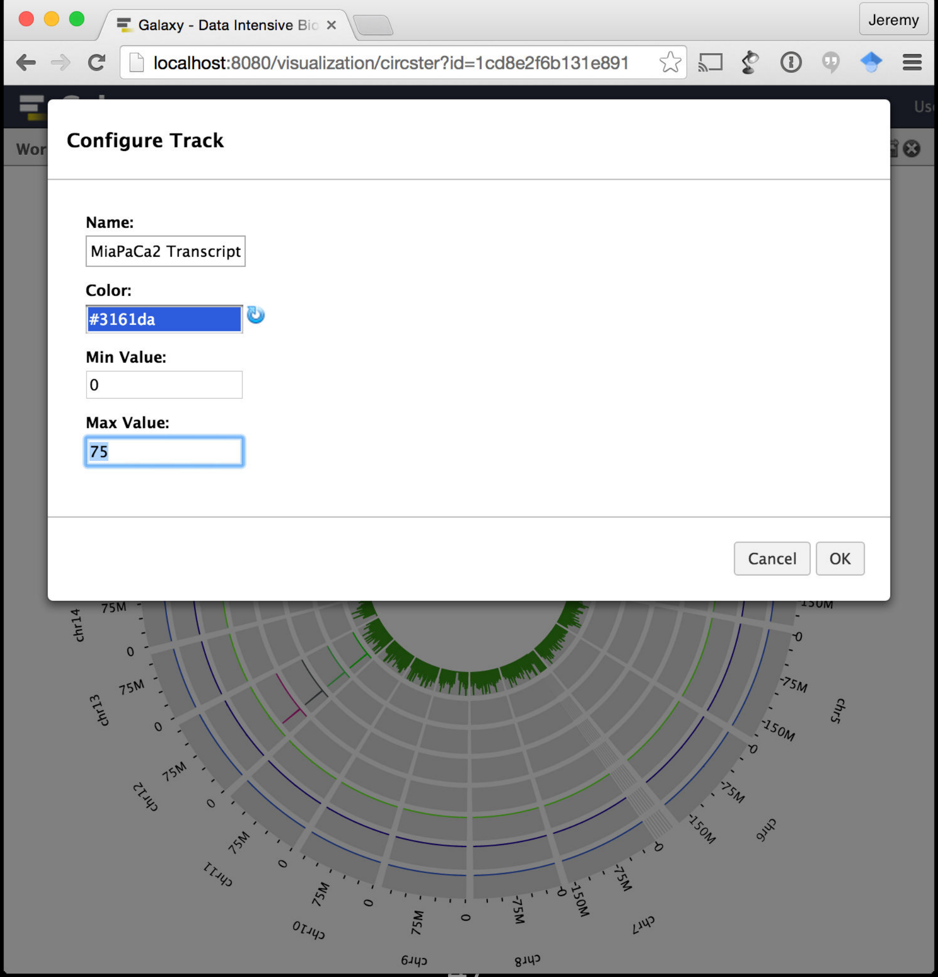
Let’s add data to Circster and adjust options

2. Change arc dataset height



Let’s add data to Circster and adjust options

3. Change max for tracks



Configure Track

Name: MiaPaCa2 Transcript

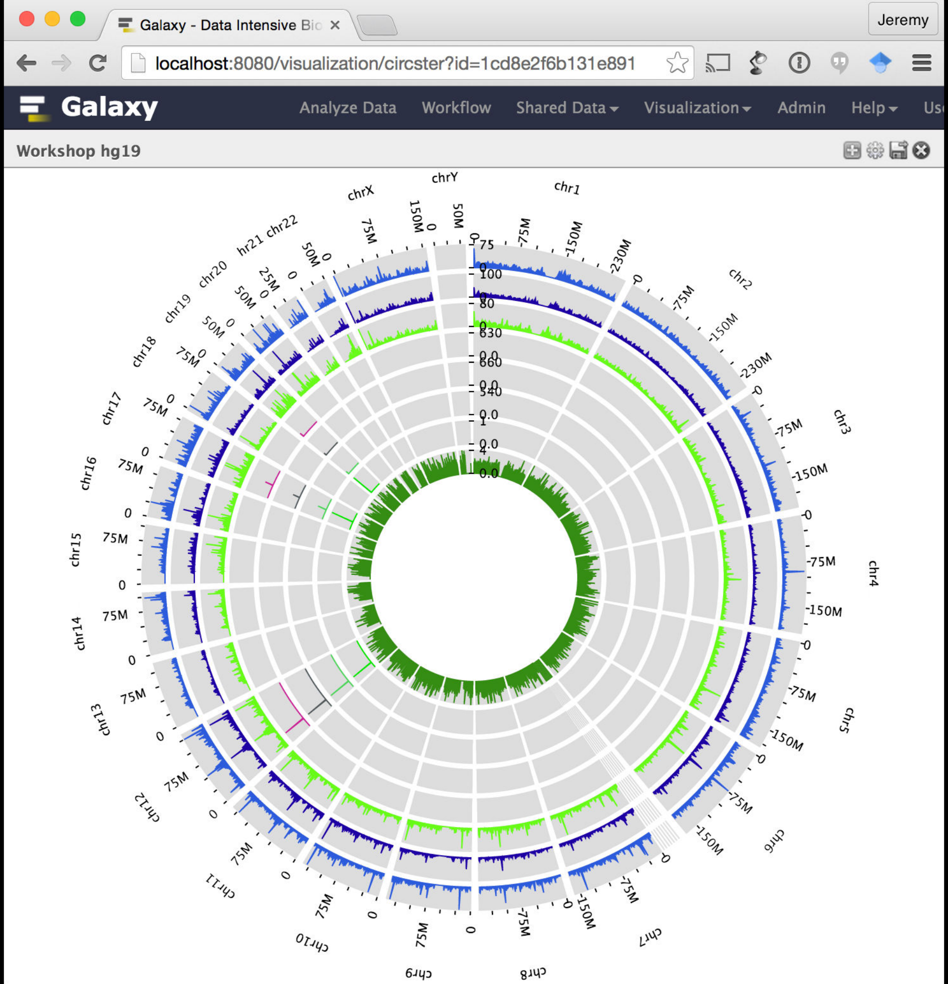
Color: #3161da

Min Value: 0

Max Value: 75

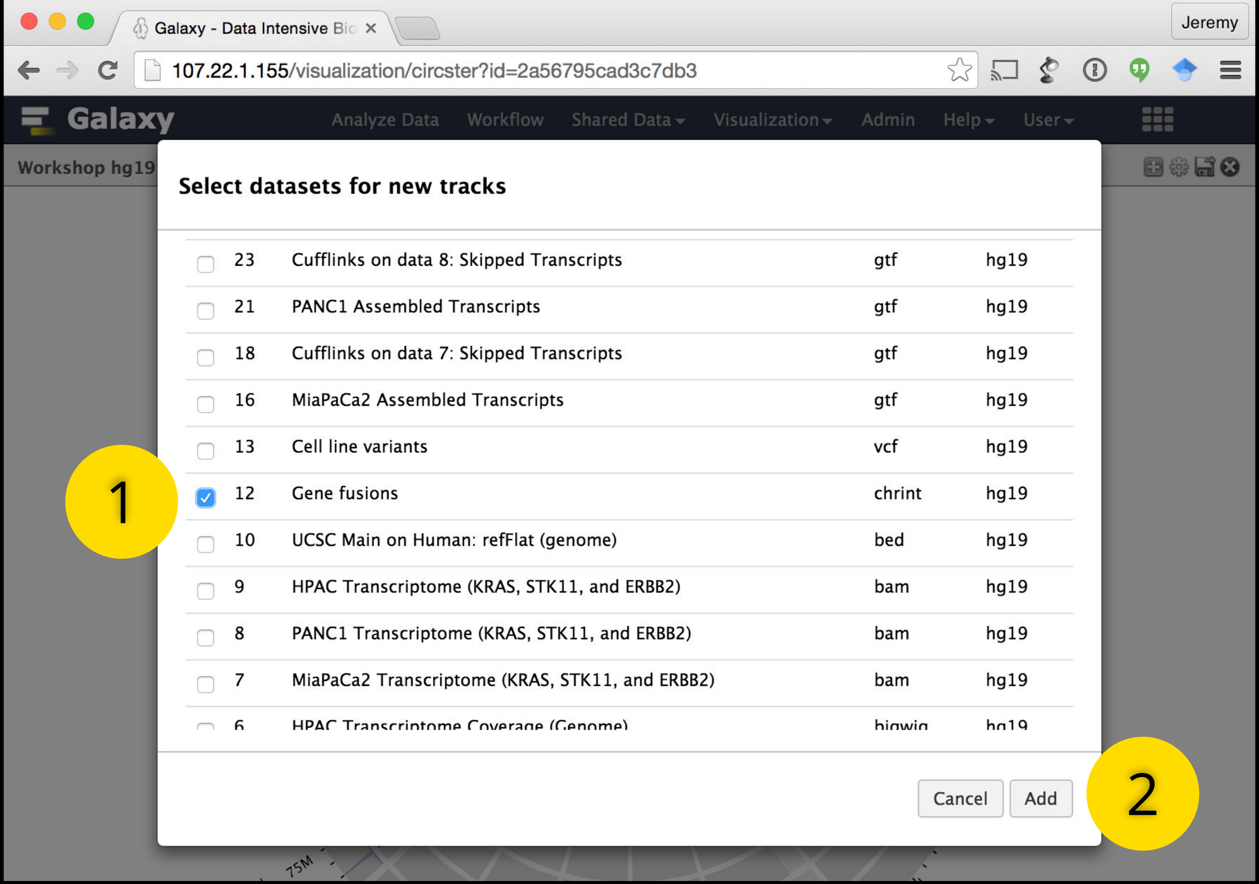
Let’s add data to Circster and adjust options

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



Let’s add data to Circster and adjust options

Add gene fusions

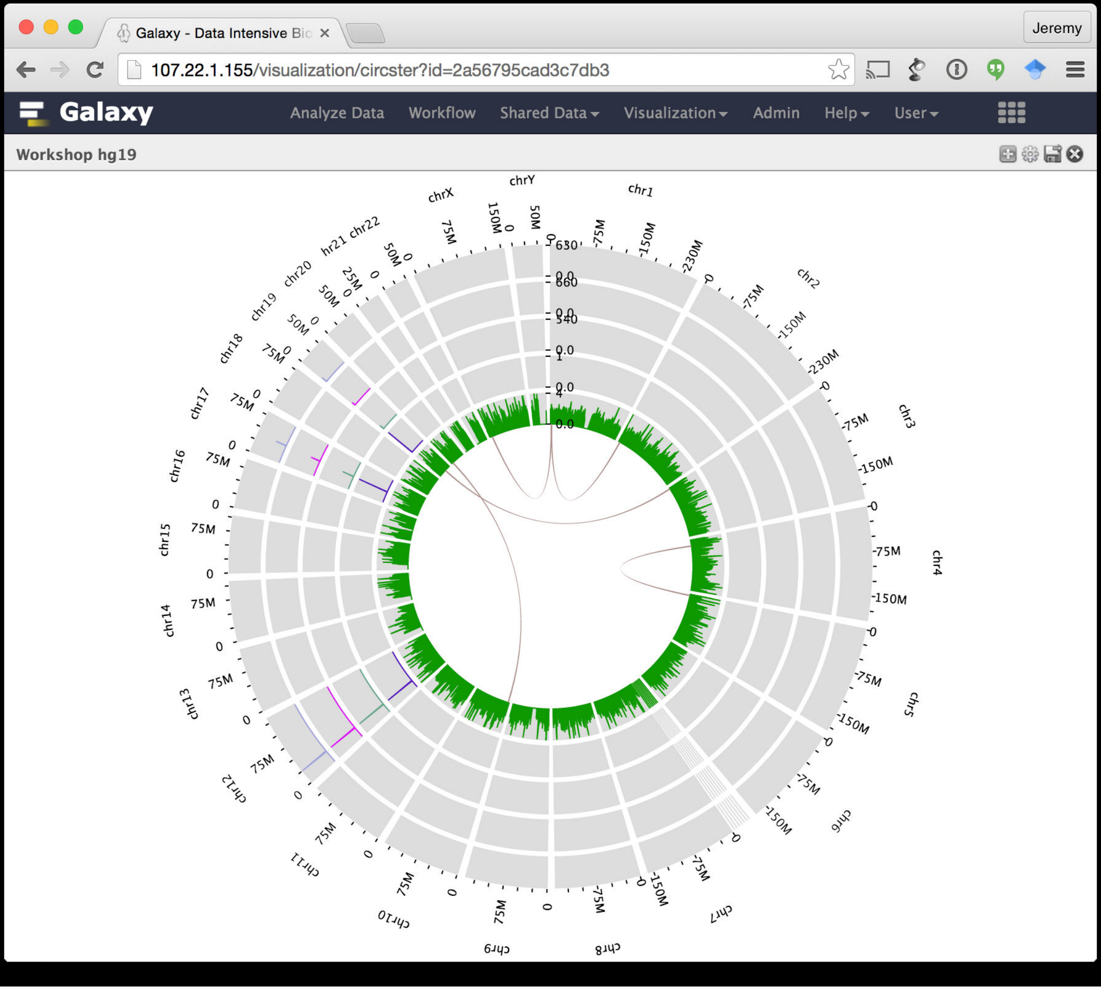


Select datasets for new tracks

☐	23	Cufflinks on data 8: Skipped Transcripts	gtf	hg19
☐	21	PANC1 Assembled Transcripts	gtf	hg19
☐	18	Cufflinks on data 7: Skipped Transcripts	gtf	hg19
☐	16	MiaPaCa2 Assembled Transcripts	gtf	hg19
☐	13	Cell line variants	vcf	hg19
☑	12	Gene fusions	chrint	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

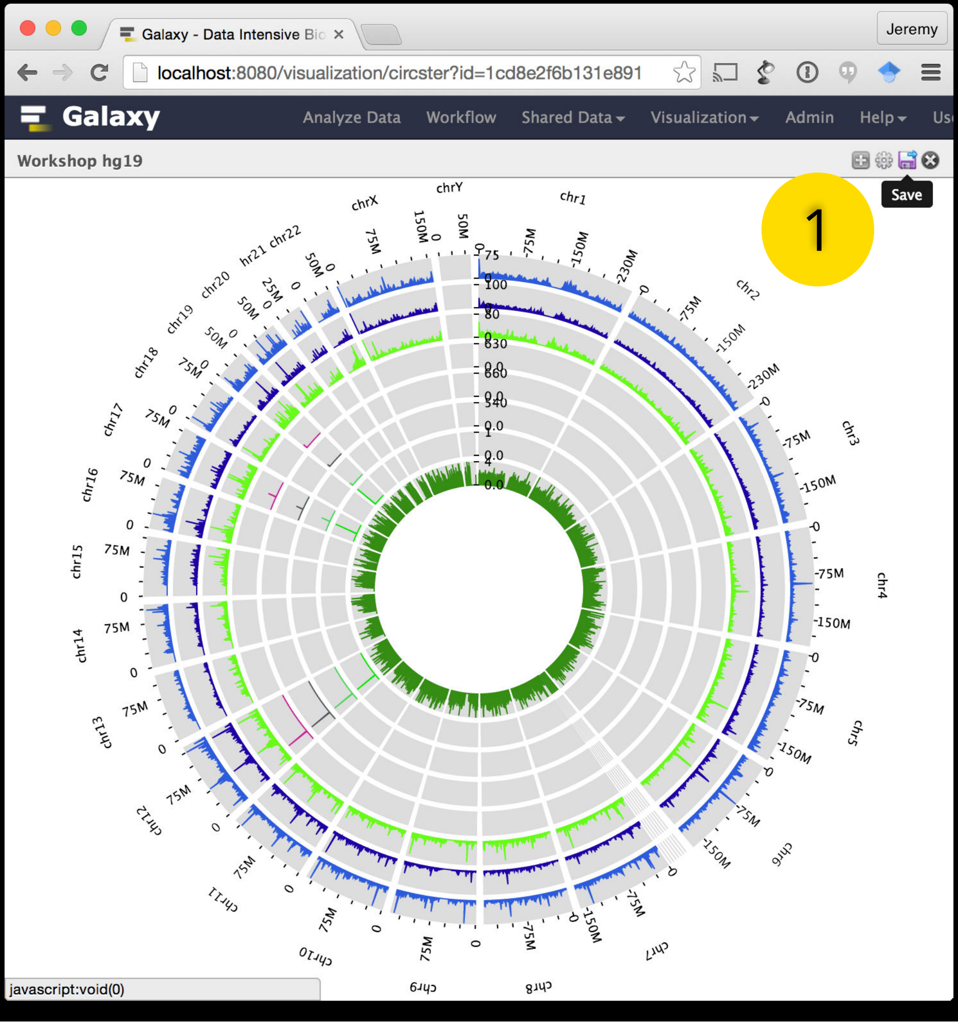
Let's add data to Circster and adjust options

Add gene fusions



Let’s add data to Circster and adjust options

4. Save visualization

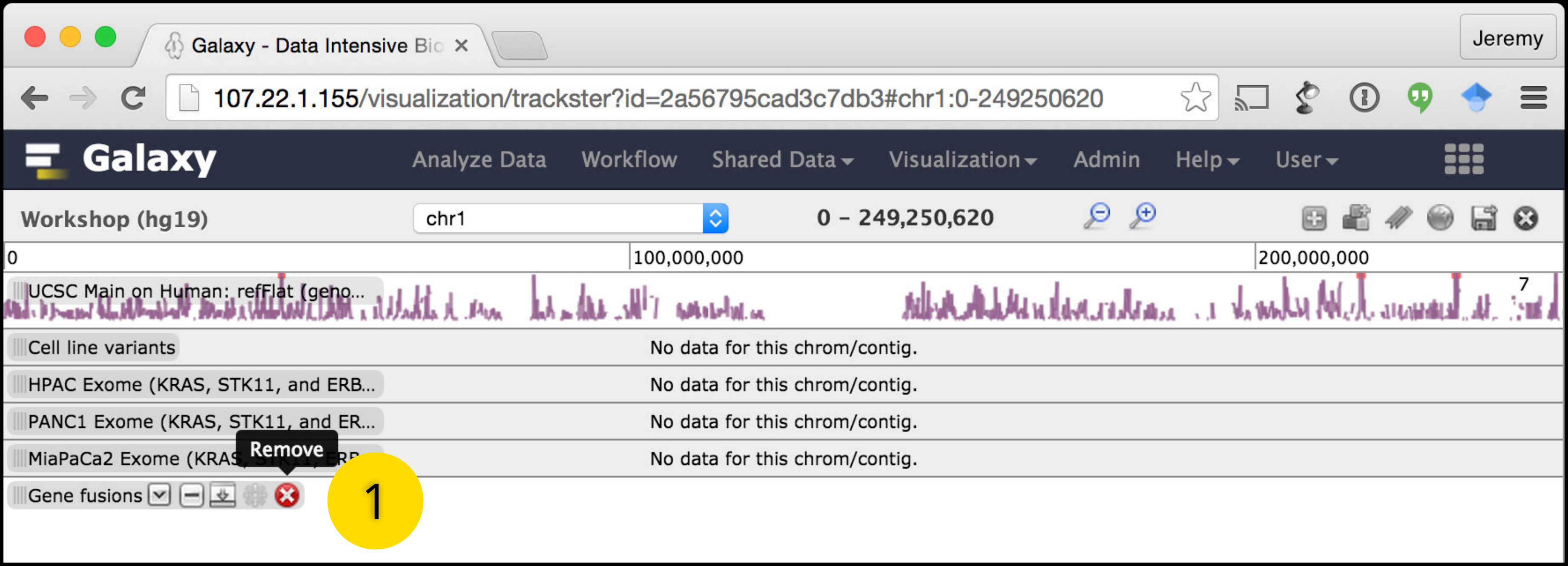


Back to Trackster: Rainbow Track for Coverage

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



No data for this chrom/contig.

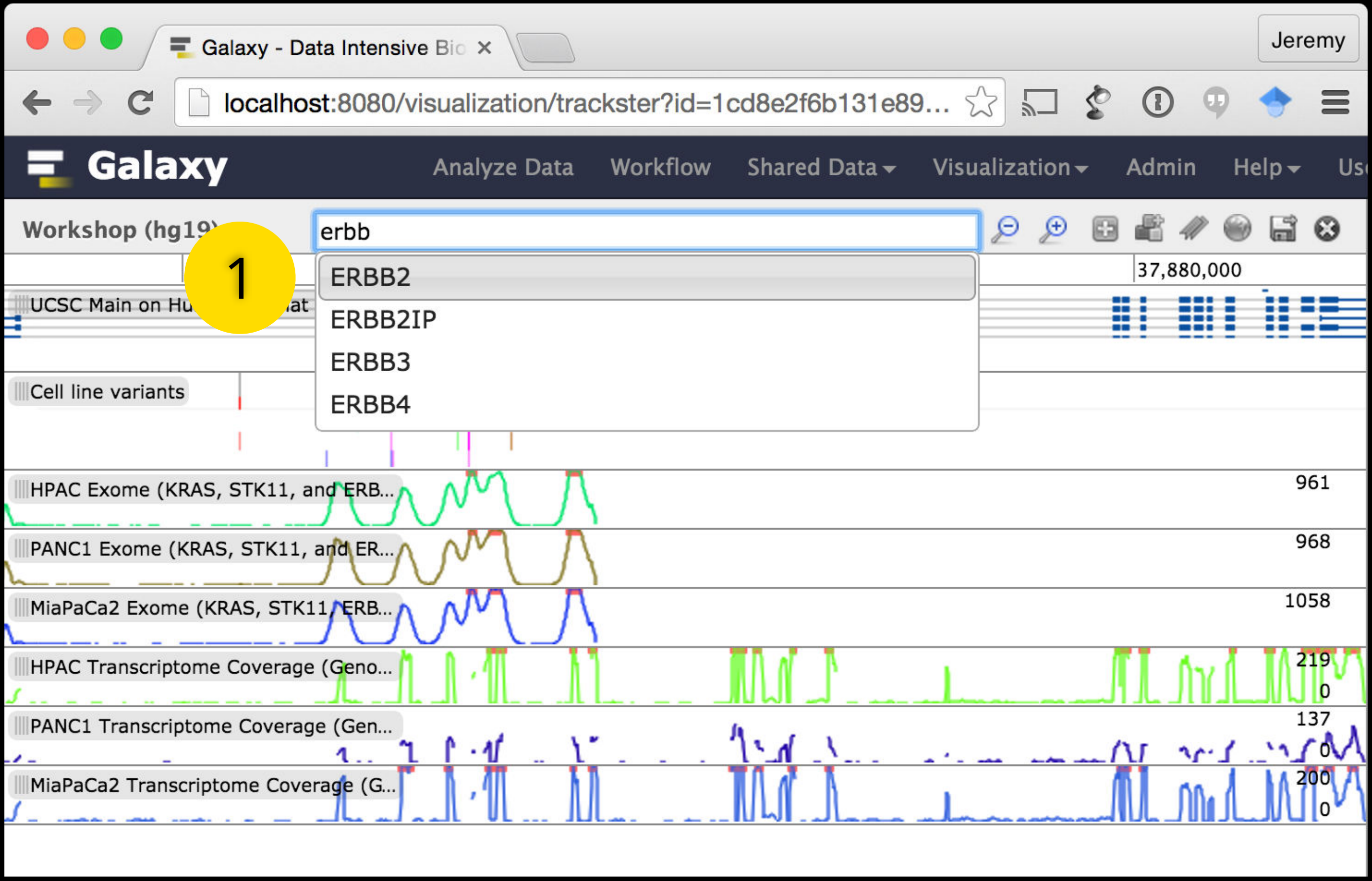
No data for this chrom/contig.

No data for this chrom/contig.

No data for this chrom/contig.

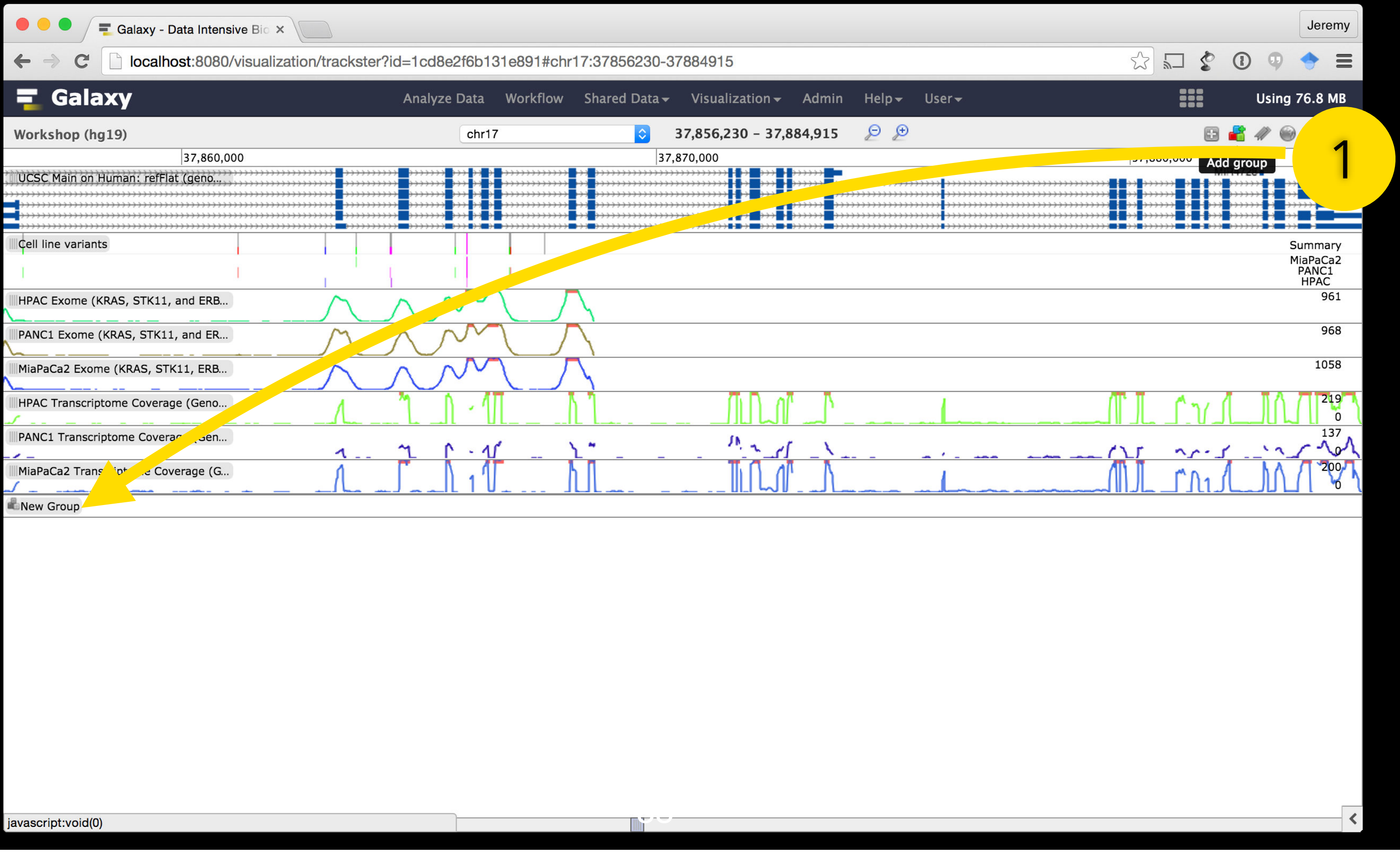
Back to Trackster: Rainbow Track for Coverage

2. Navigate to ERBB2 gene



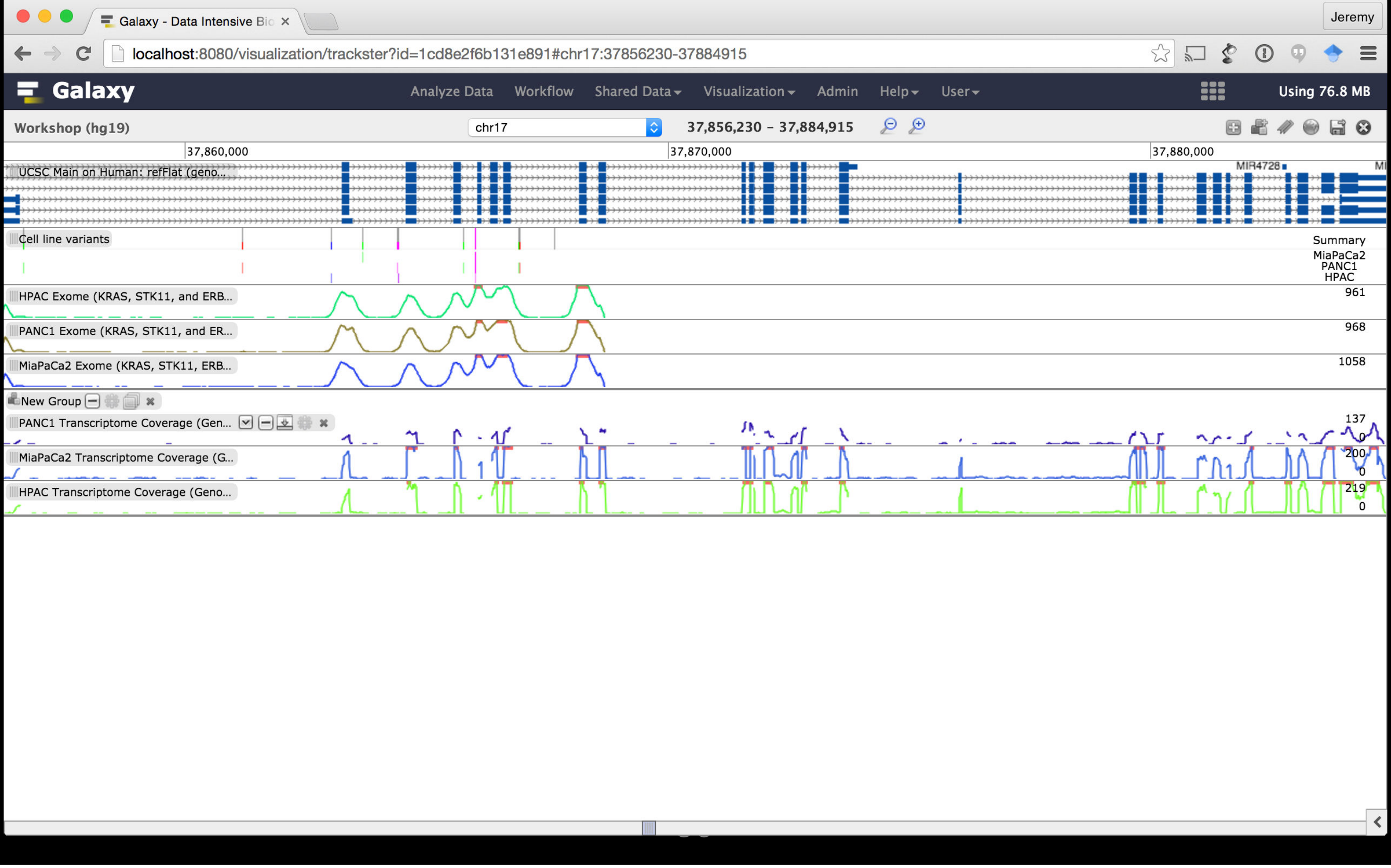
Back to Trackster: Rainbow Track for Coverage

3. Create group



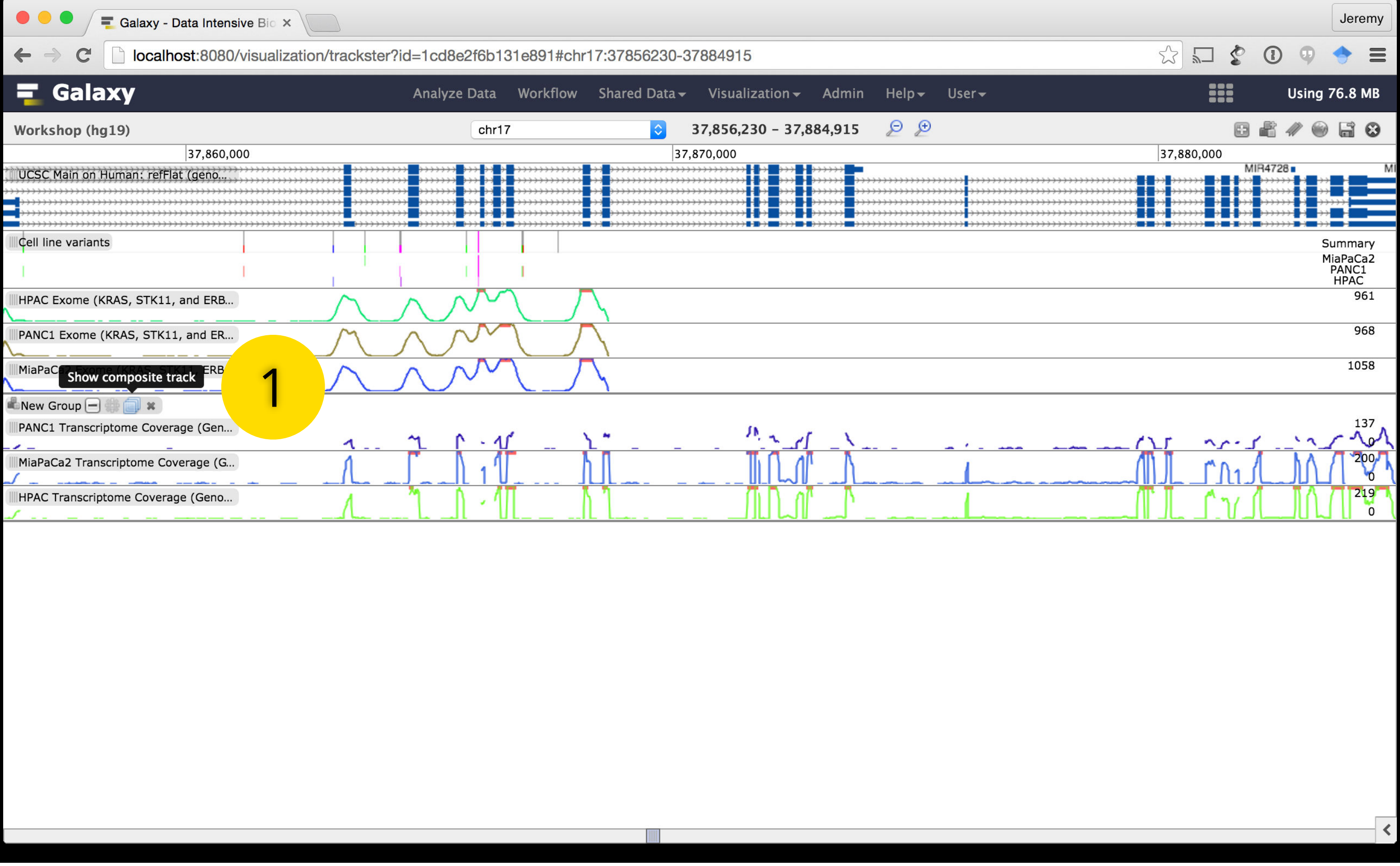
Back to Trackster: Rainbow Track for Coverage

4. Add transcriptome coverage tracks to group



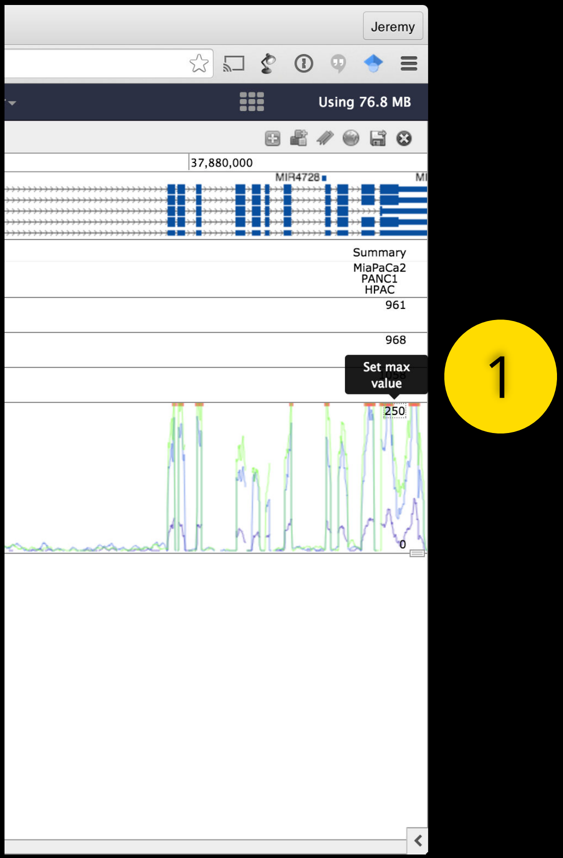
Back to Trackster: Rainbow Track for Coverage

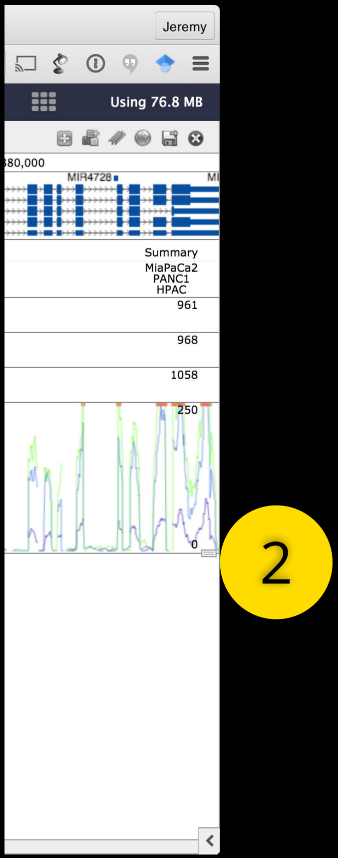
5. Create composite track



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

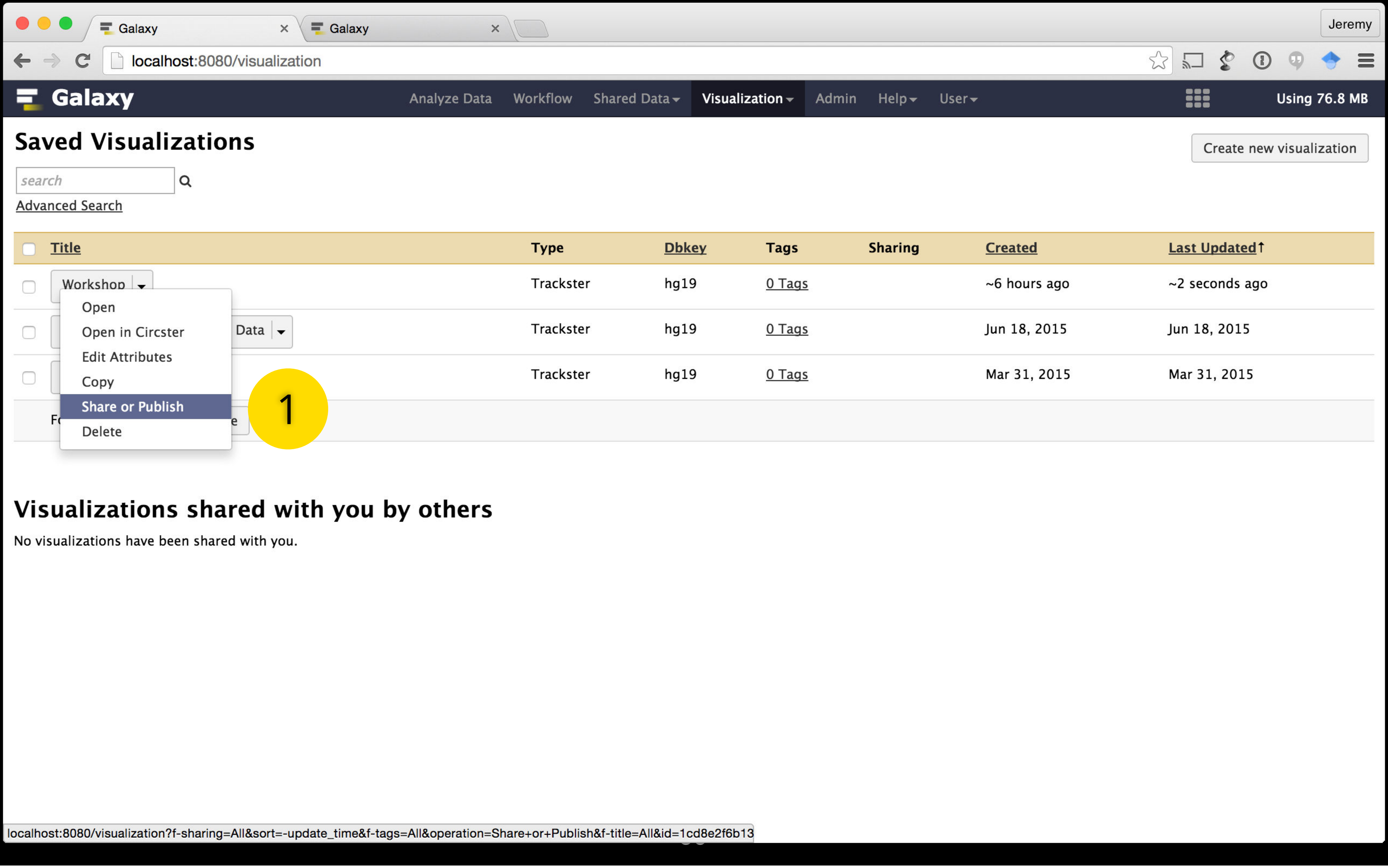
Look at STK11

- bookmark

Look at KRAS

- bookmark

Share and Publish



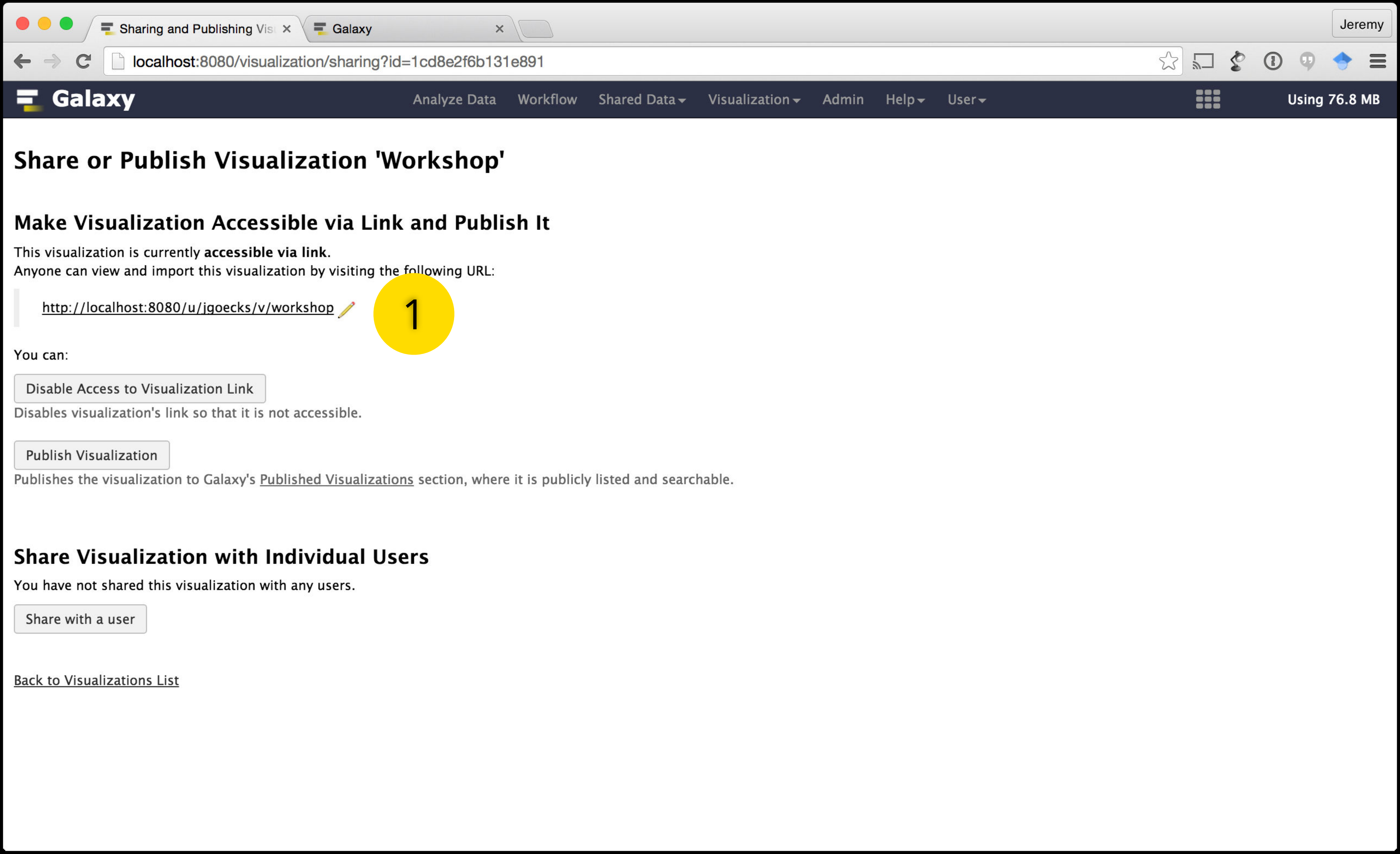
Saved Visualizations

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

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



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

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.

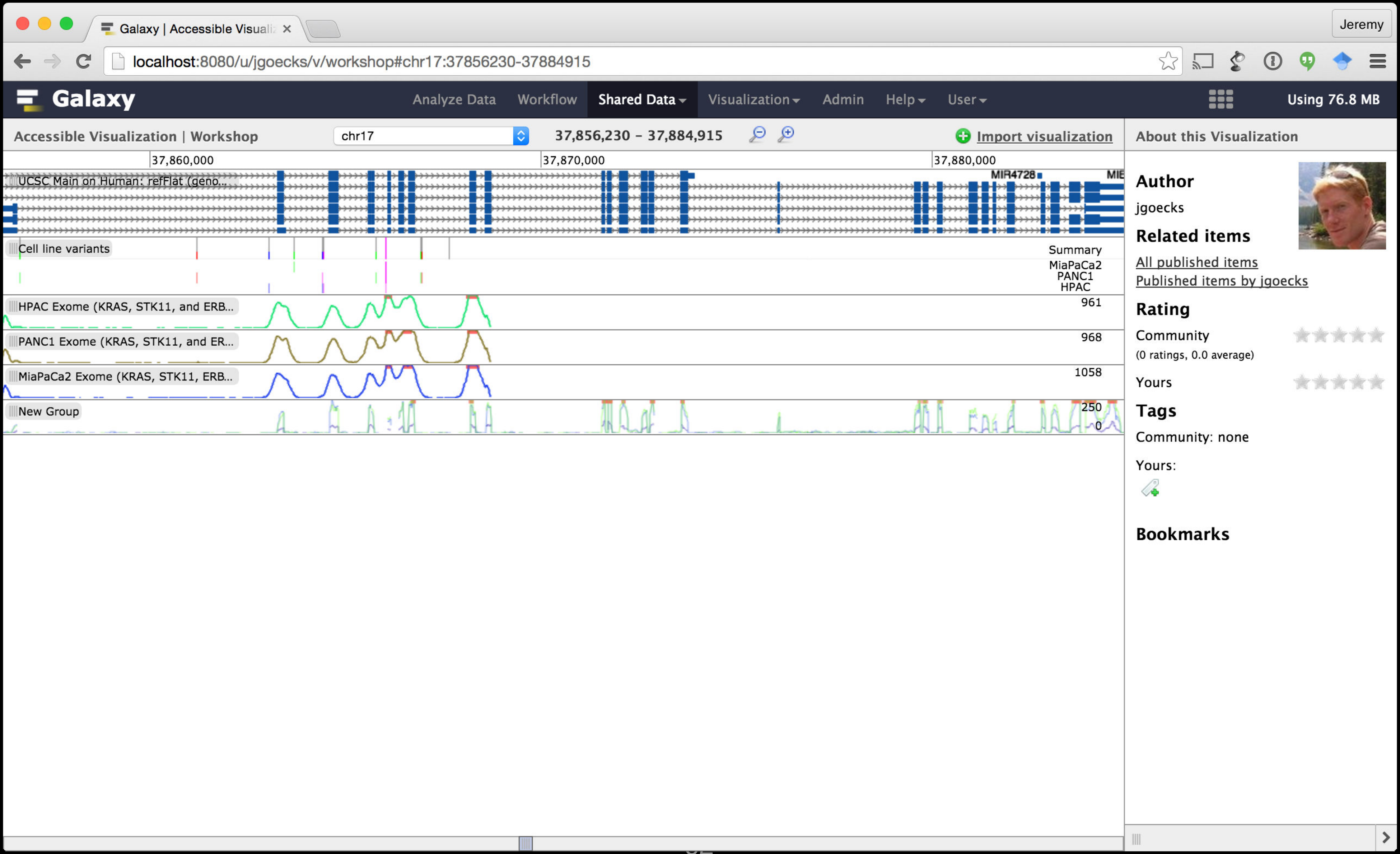
Share Visualization with Individual Users

You have not shared this visualization with any users.

Share with a user

Back to Visualizations List

Share and Publish



Demo: Visual Analysis

Topics

Visualization history and introduction

Numerical Visualizations

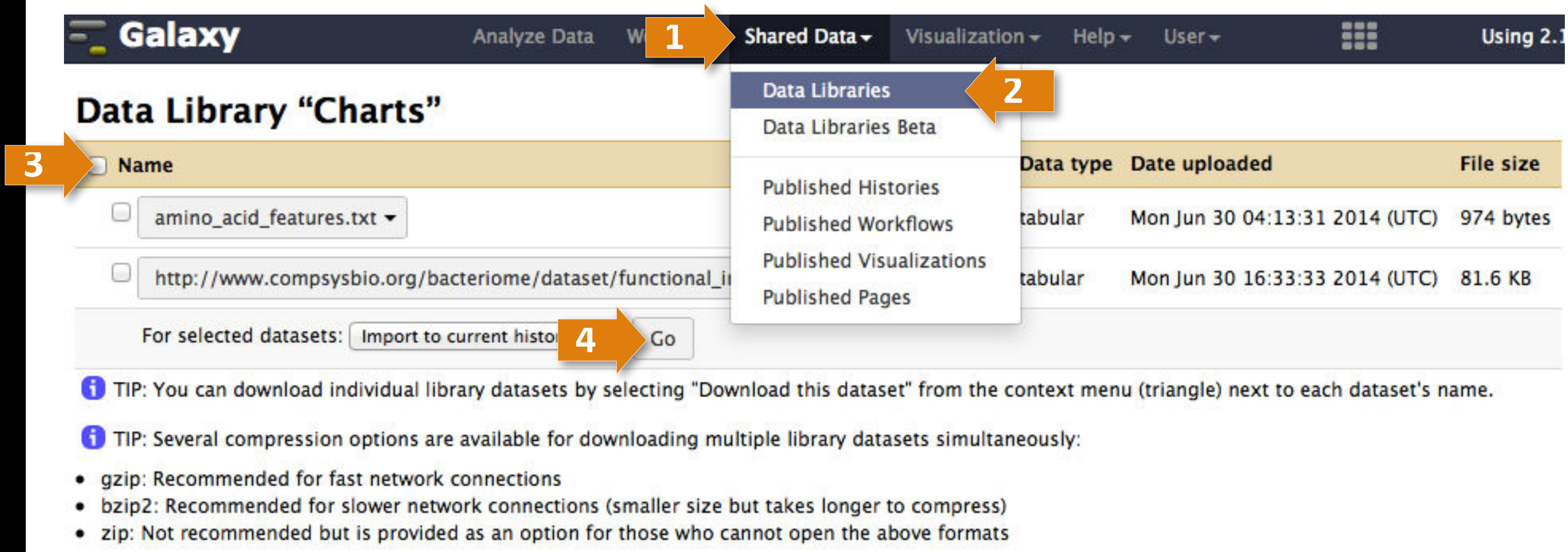
Biological Visualizations

Adding your own visualizations

What is Galaxy Charts?

Use Galaxy | **Create Tabular Results** | **Visualize with Galaxy *Charts***

Import data files



Data Library “Charts”

☐ Name	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_i	tabular	Mon Jun 30 16:33:33 2014 (UTC)	81.6 KB

For selected datasets: Import to current histo

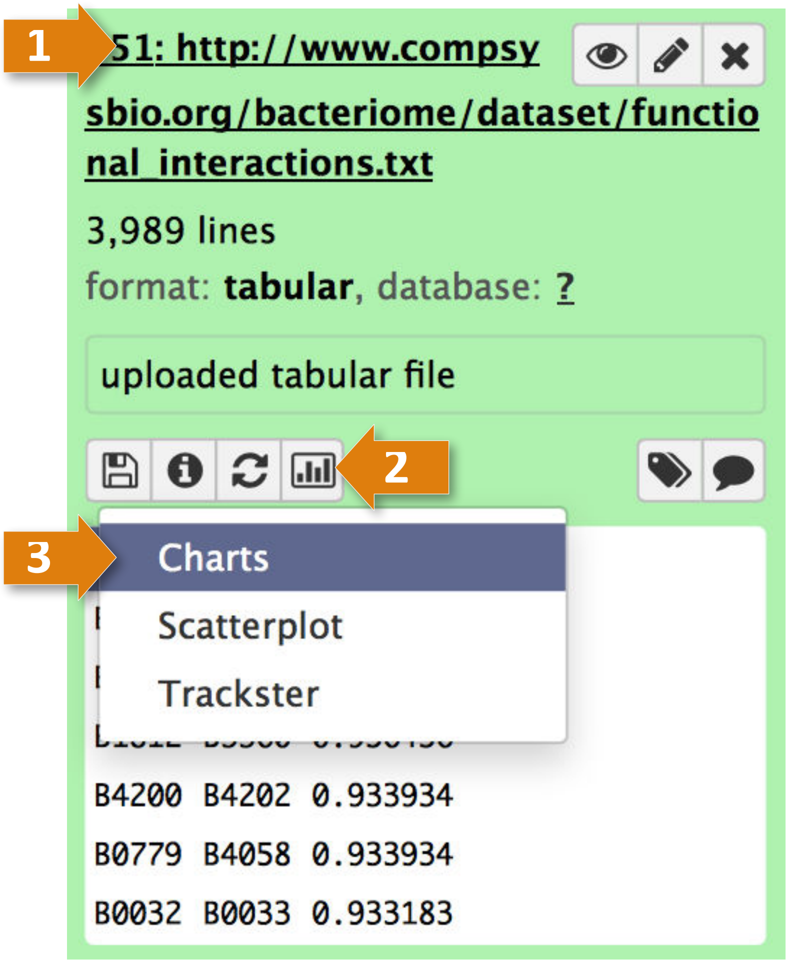
TIP: You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.

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

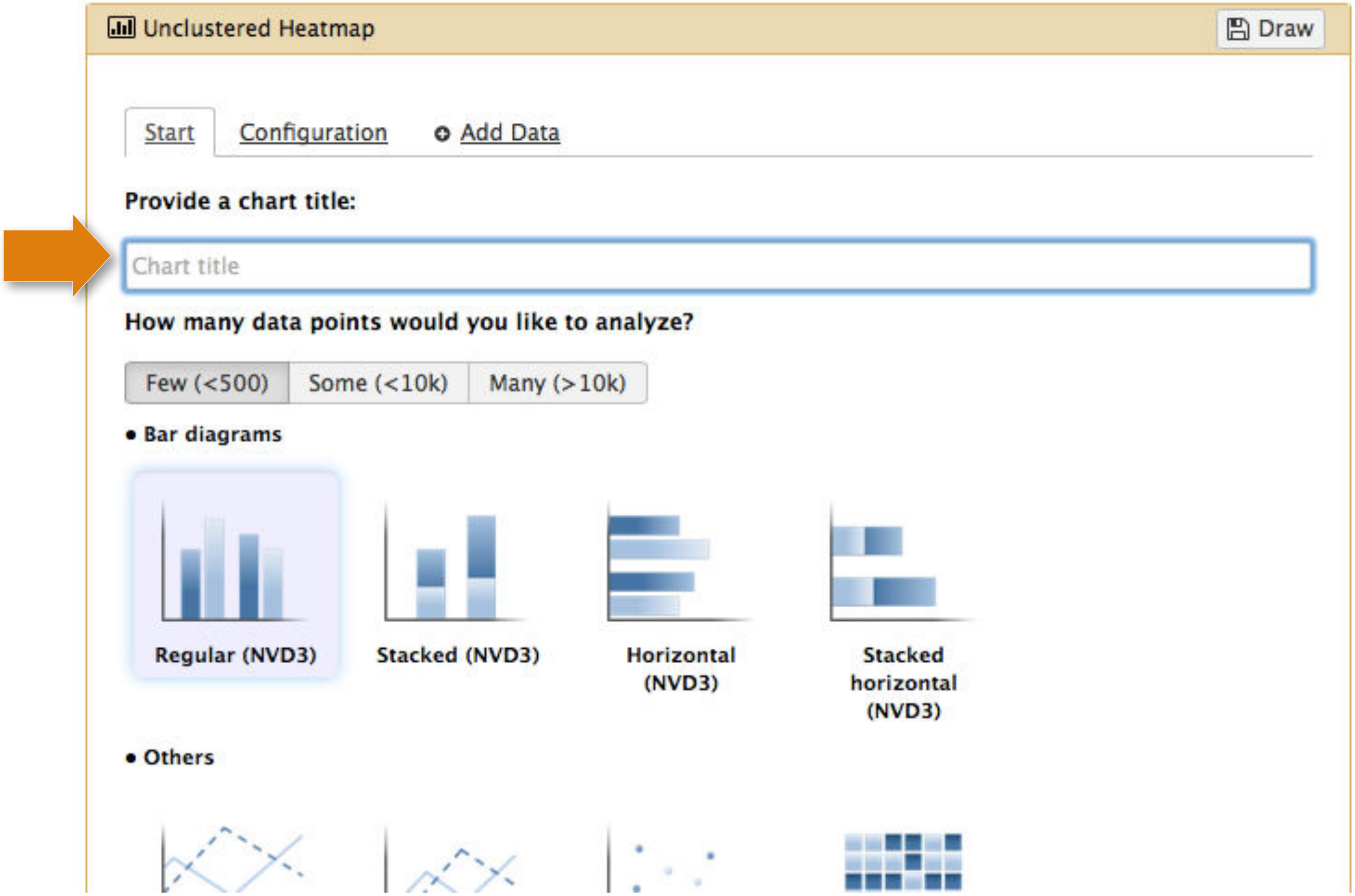
Click on **Shared Data** and select **Data Libraries**. Navigate to the **Chart** library and import it into your history (*data reference: http://dna.cs.byu.edu/treesaap and bacteriome.org*).

Make a new chart (1 of 4)



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

Give your chart a name



Provide a chart title:

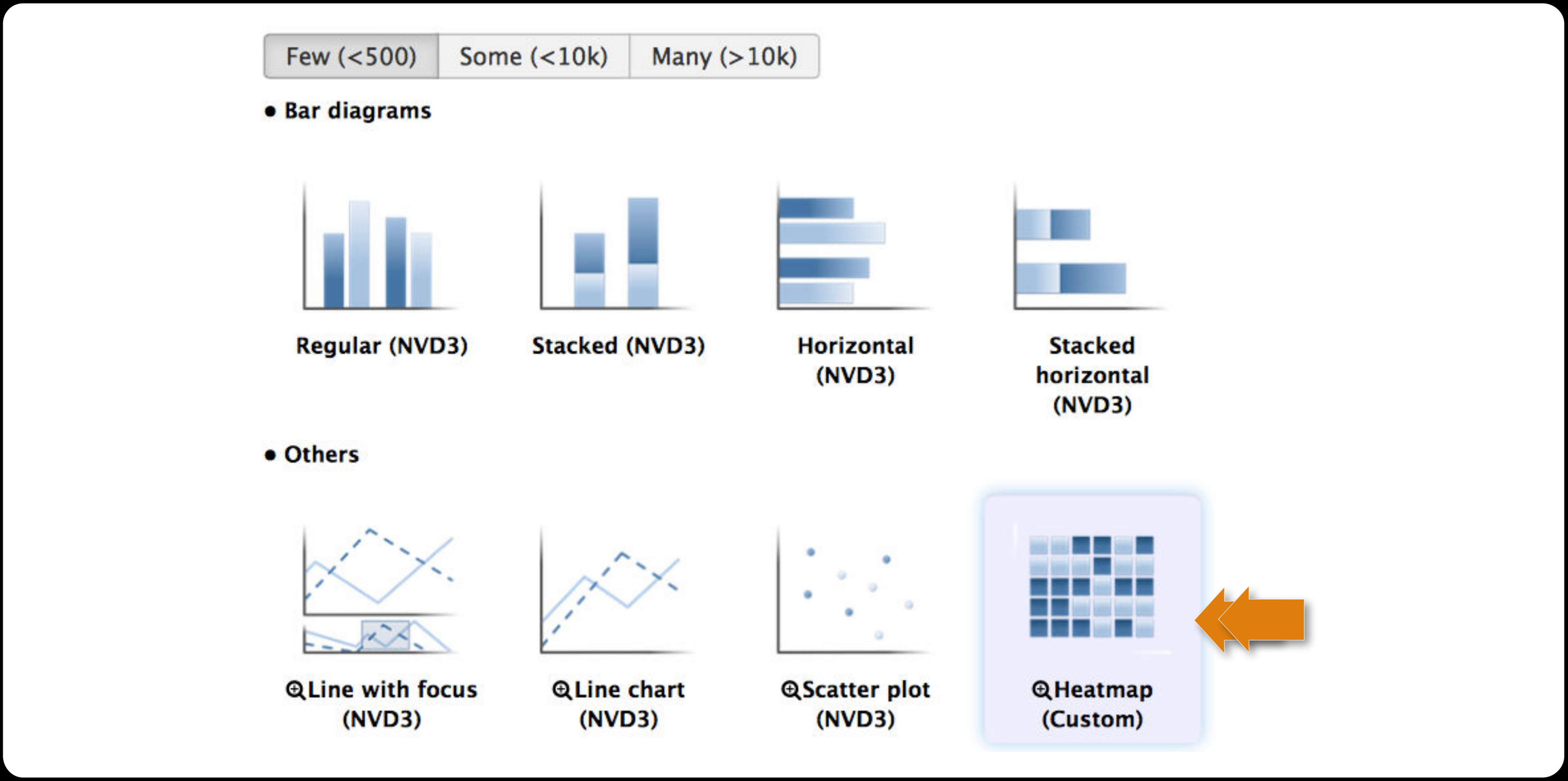
Chart title

How many data points would you like to analyze?

- Bar diagrams
- Others

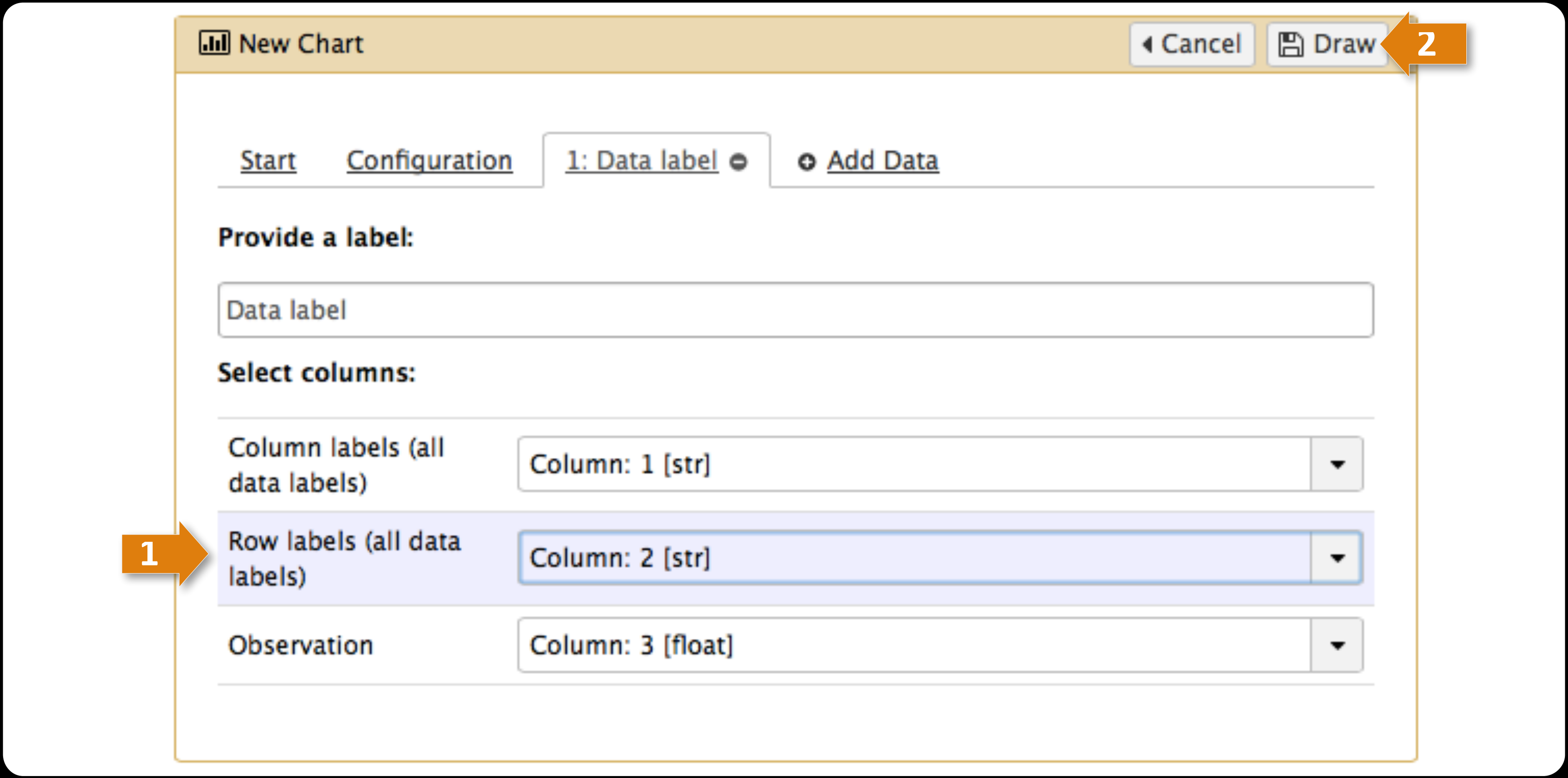
Name your chart **Unclustered Heatmap**.

Select a chart type



Double click on the **Heatmap** icon.

Select data columns



Provide a label:

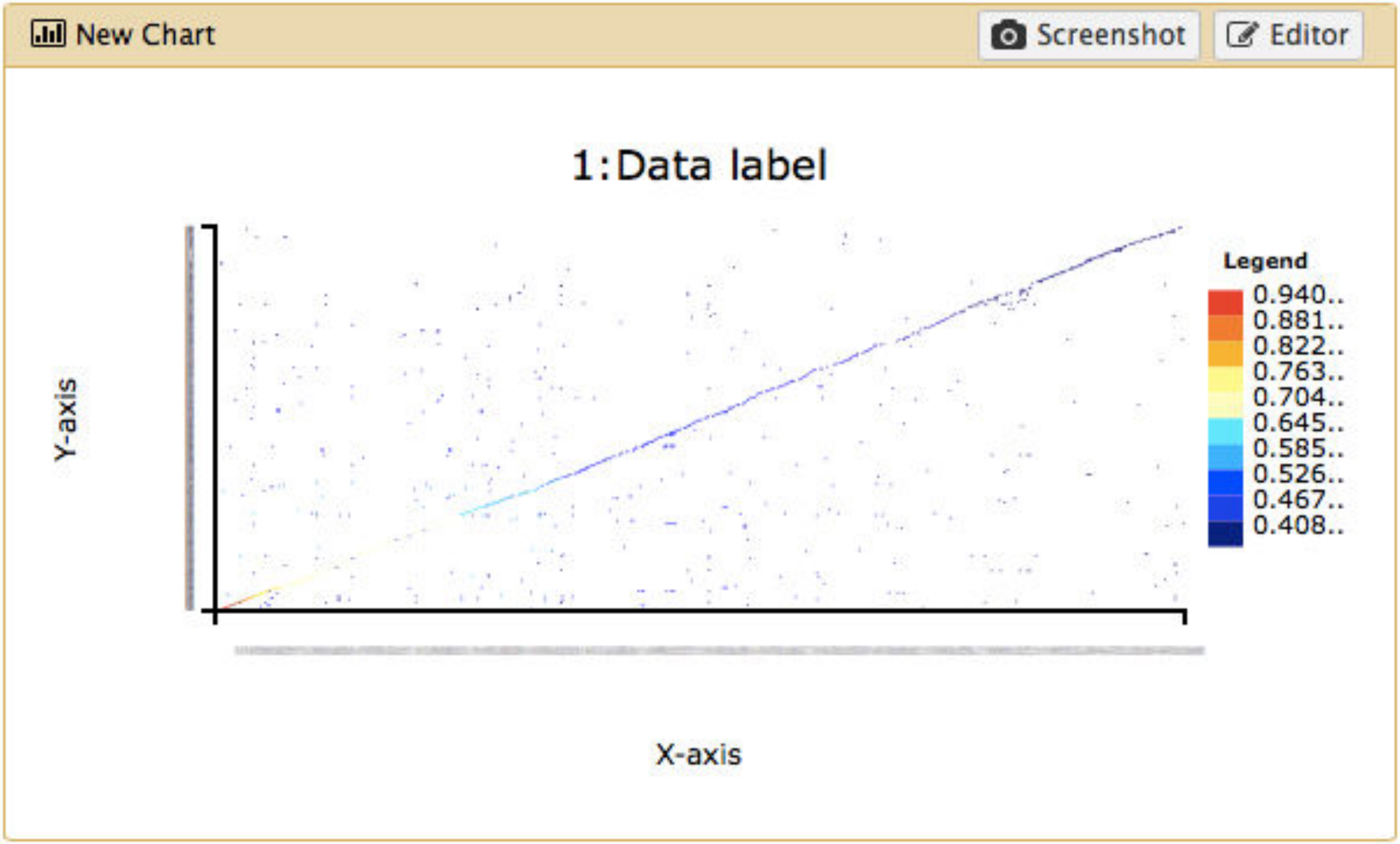
Data label

Select columns:

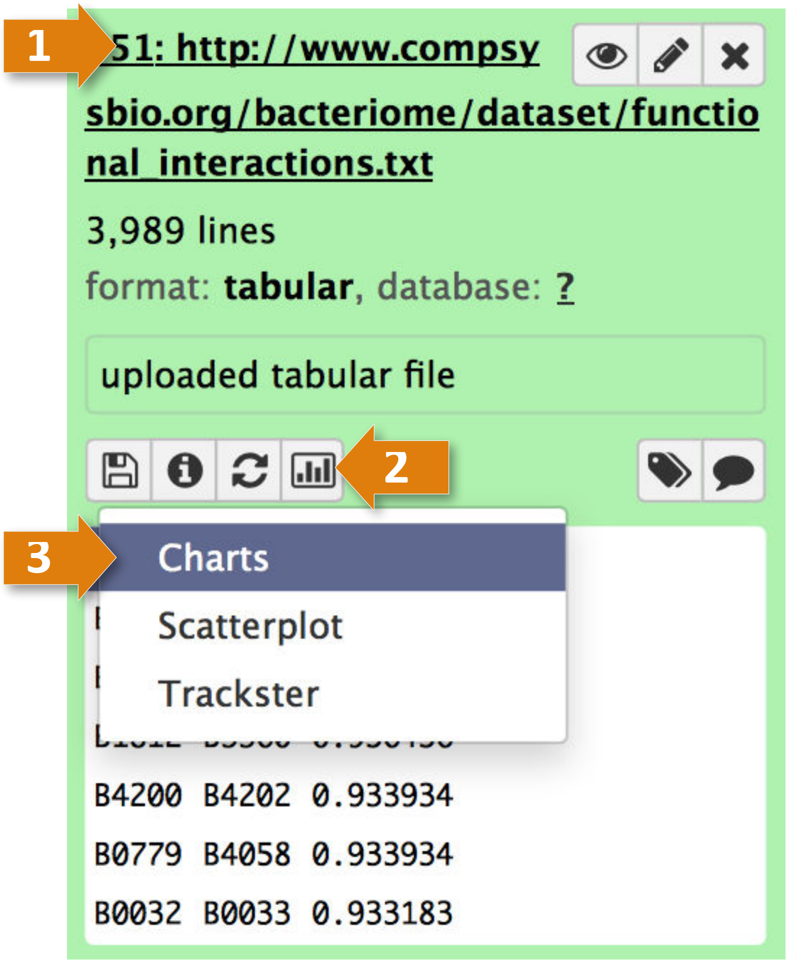
Column labels (all data labels)	Column: 1 [str]
Row labels (all data labels)	Column: 2 [str]
Observation	Column: 3 [float]

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

Unclustered Heatmap



Make a new chart (2 of 4)



51: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

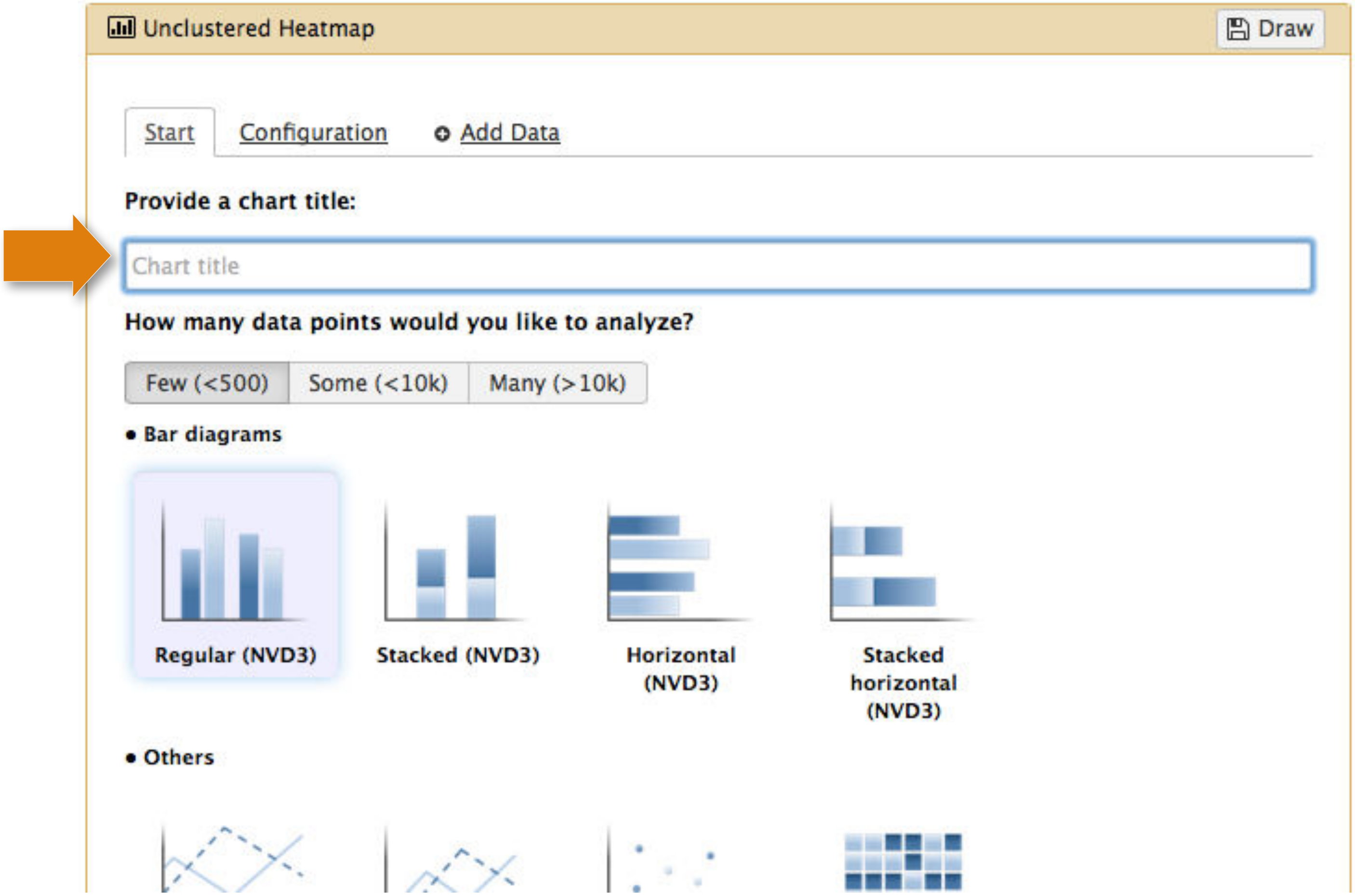
format: **tabular**, database: ?

uploaded tabular file

```
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



Provide a chart title:

Chart title

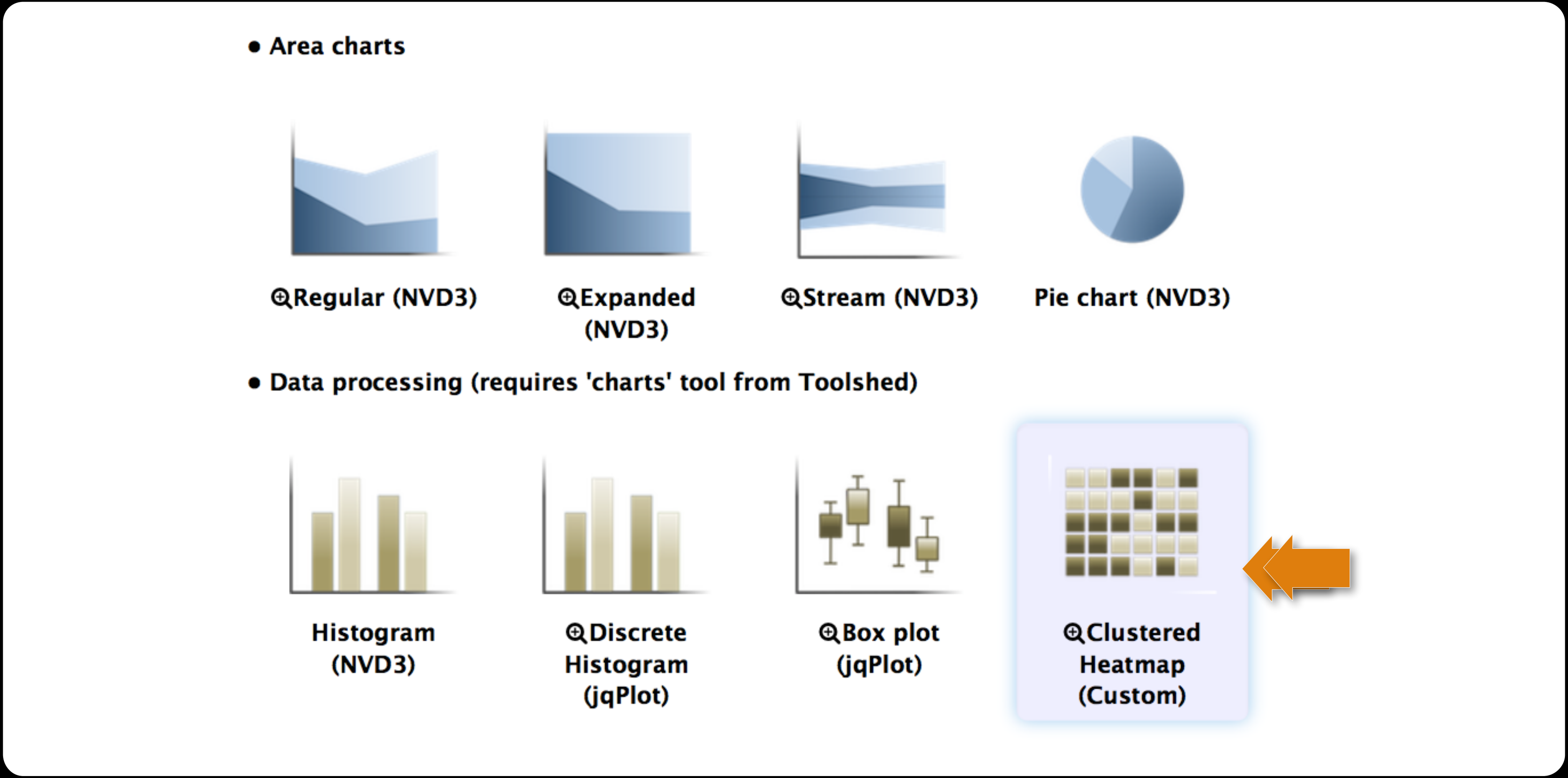
How many data points would you like to analyze?

- **Bar diagrams**

- **Others**

Name your chart **Clustered Heatmap**.

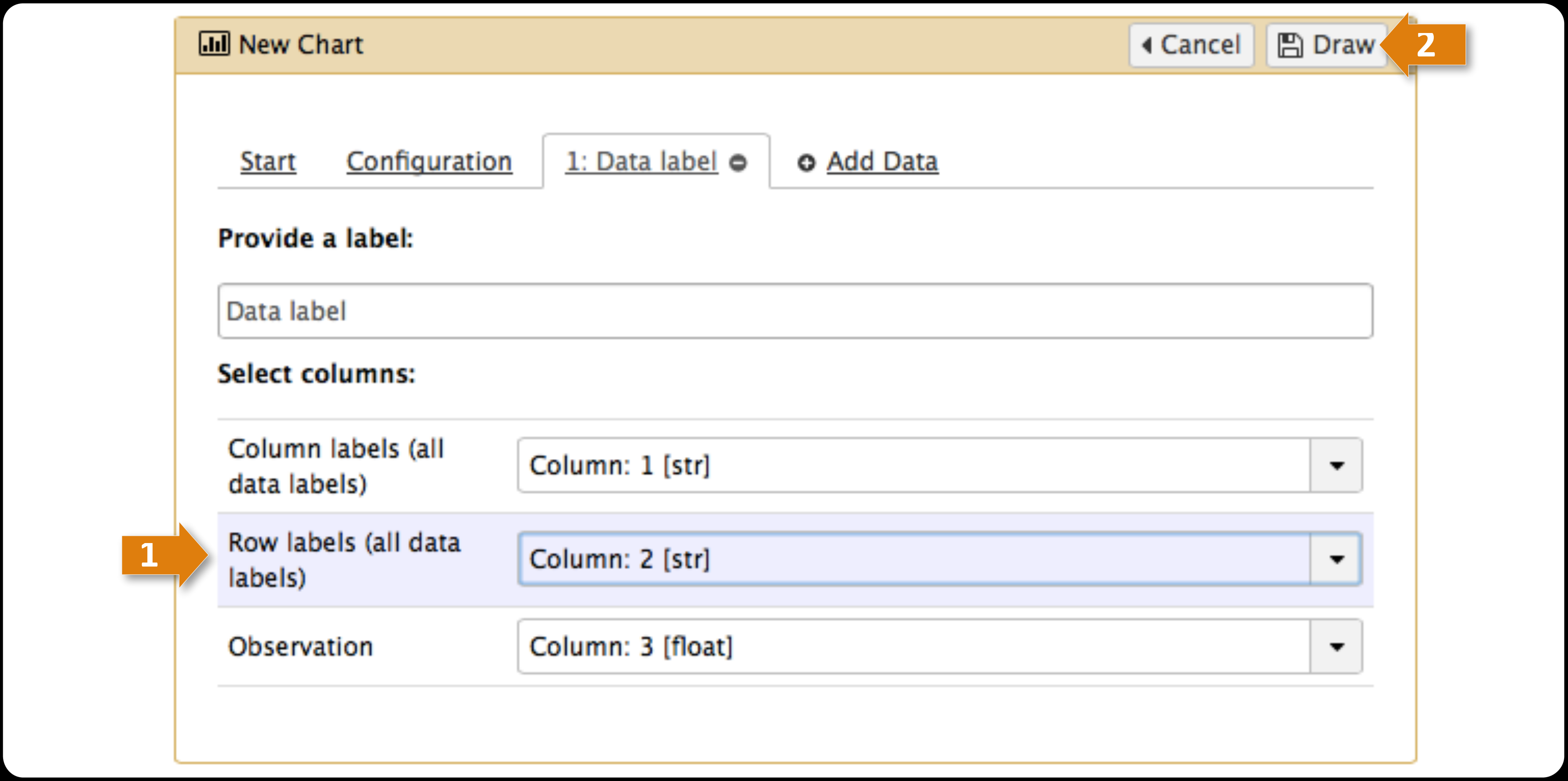
Select a new chart type



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

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

Select data columns



Provide a label:

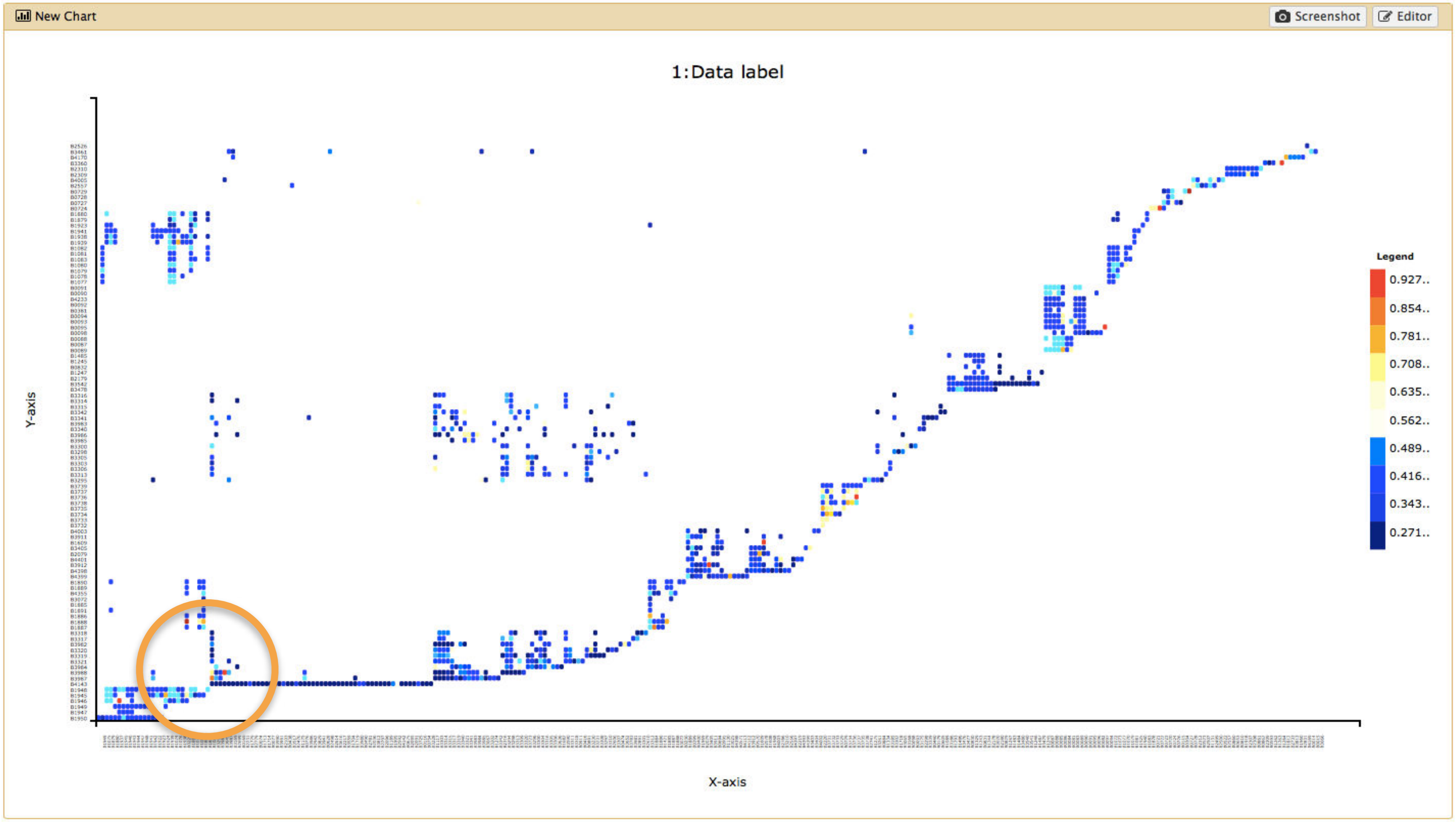
Data label

Select columns:

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

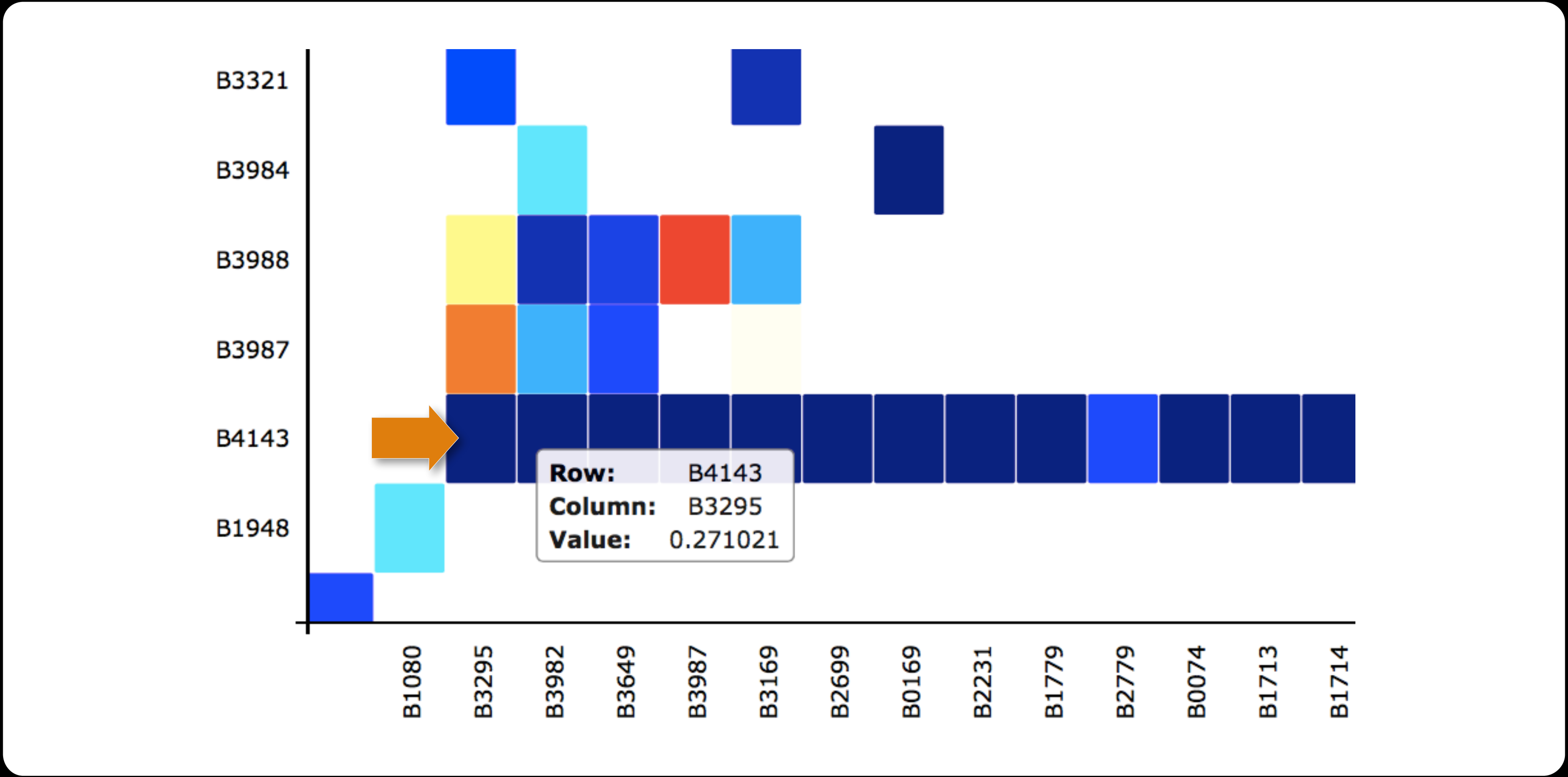
At first click on **Row labels** and select **Column 2**. Then, click on **Draw**.

Clustered Heatmap



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

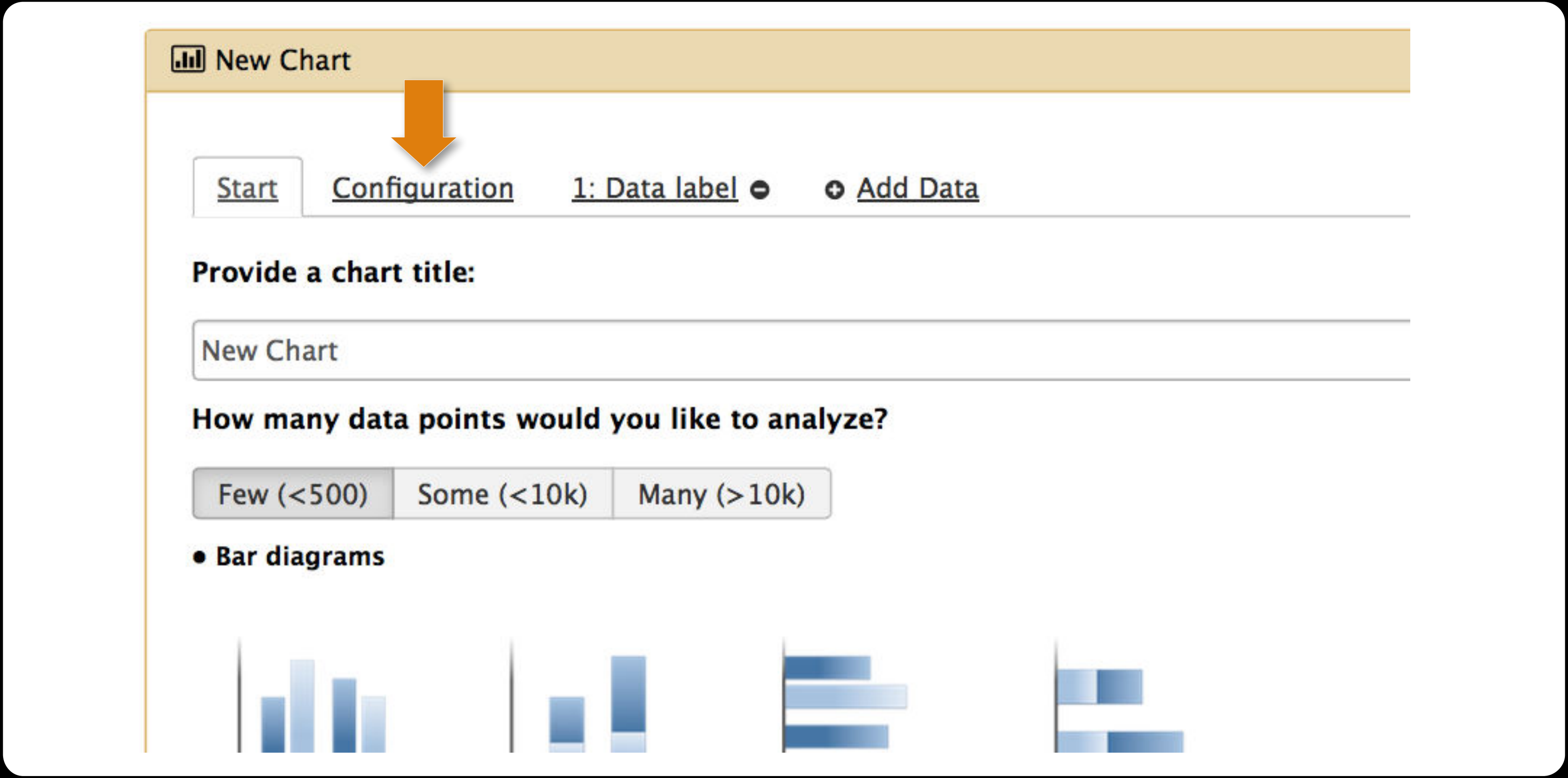
Enlarged view



Row: B4143
Column: B3295
Value: 0.271021

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

Provide a chart title:

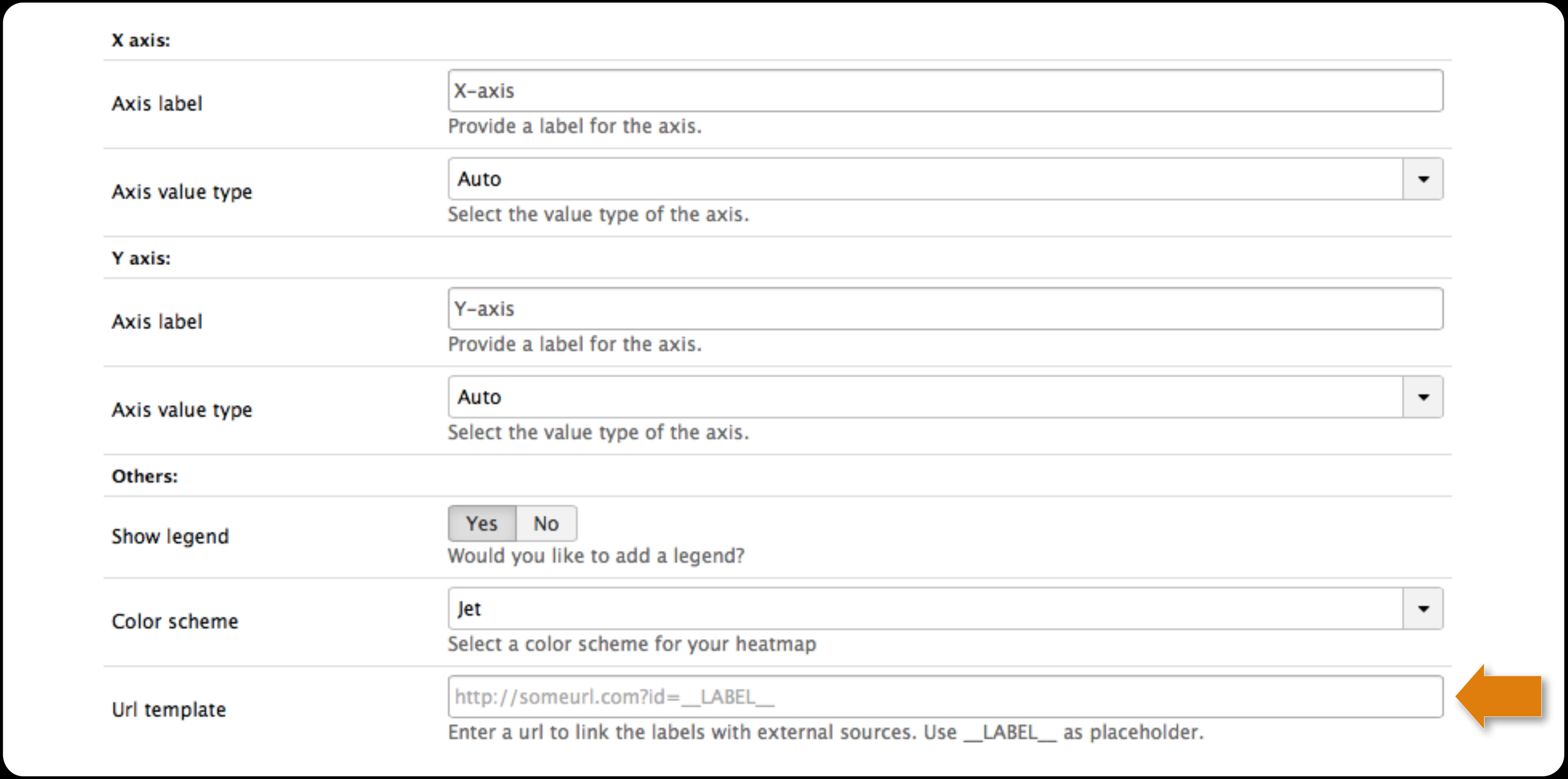
New Chart

How many data points would you like to analyze?

- **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:

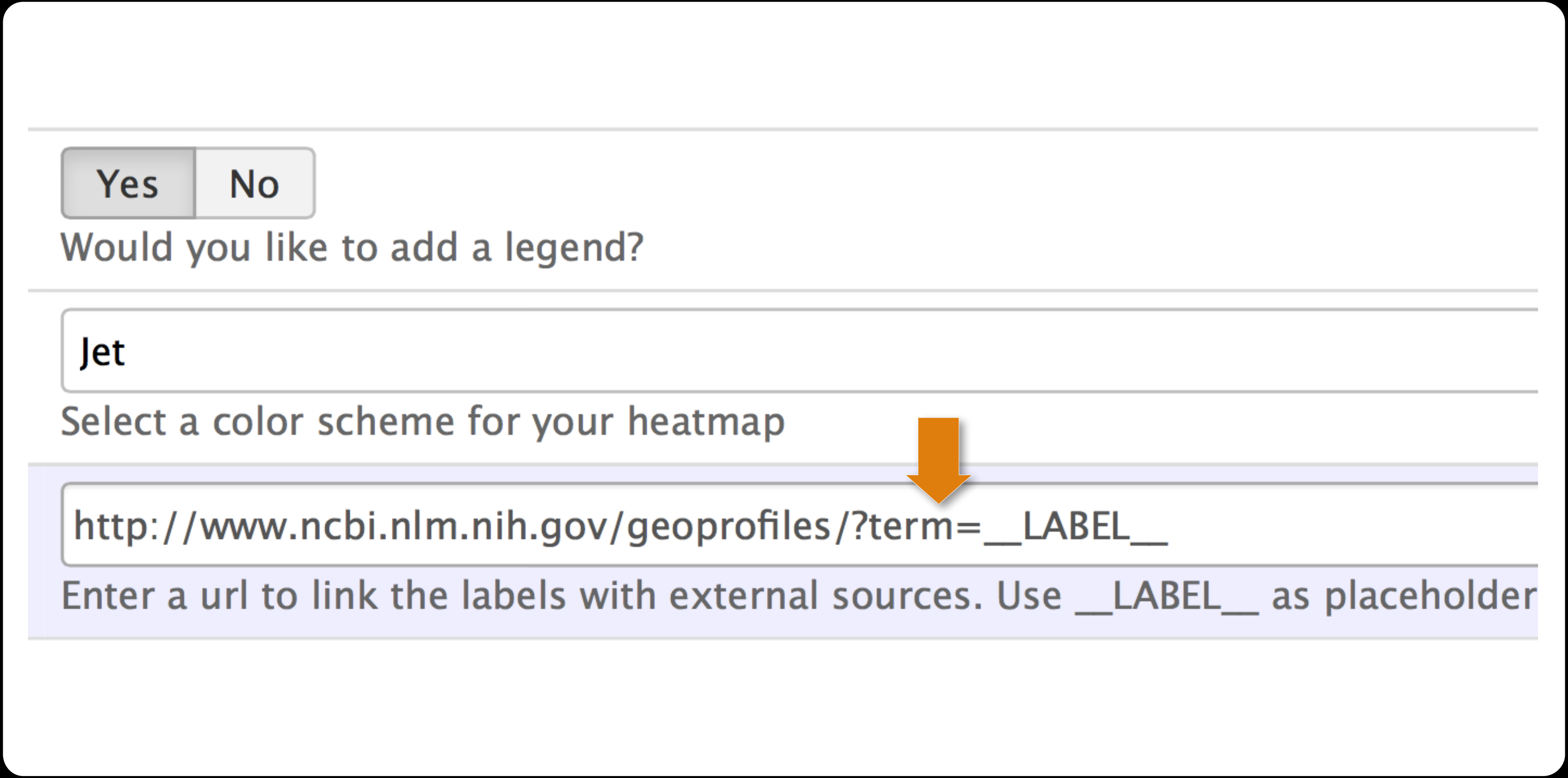
Show legend: [✓ Yes] No
Would you like to add a legend?

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



Would you like to add a legend?

Jet

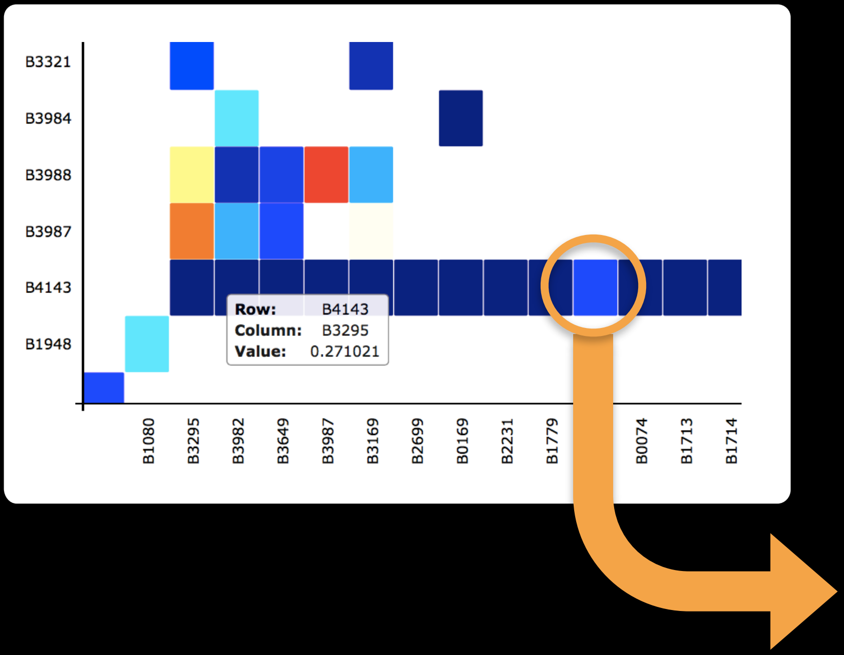
Select a color scheme for your heatmap

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

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

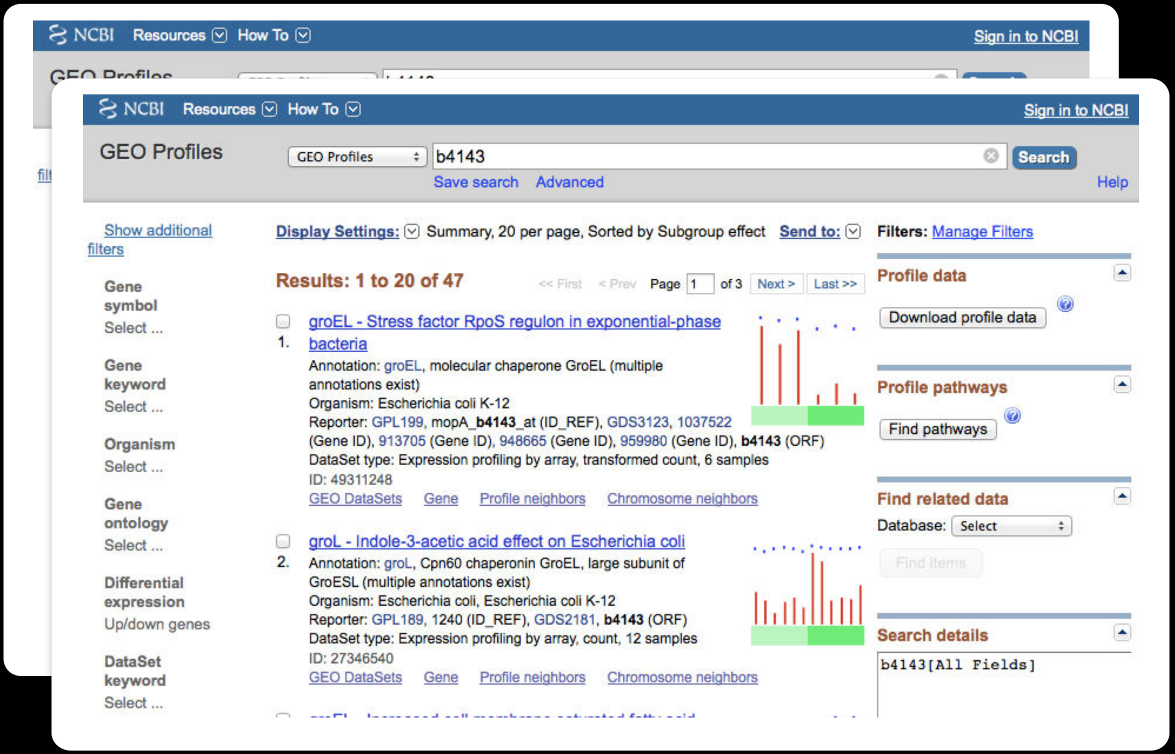
Paste a **database URL** into the template URL field and add the **__LABEL__** tag. You may use **http://www.ncbi.nlm.nih.gov** or any other database. Click on **Draw** to redraw the chart.

Data points linked to web sources



Row: B4143
Column: B3295
Value: 0.271021

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



b4143

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

Results: 1 to 20 of 47

Page 1 of 3

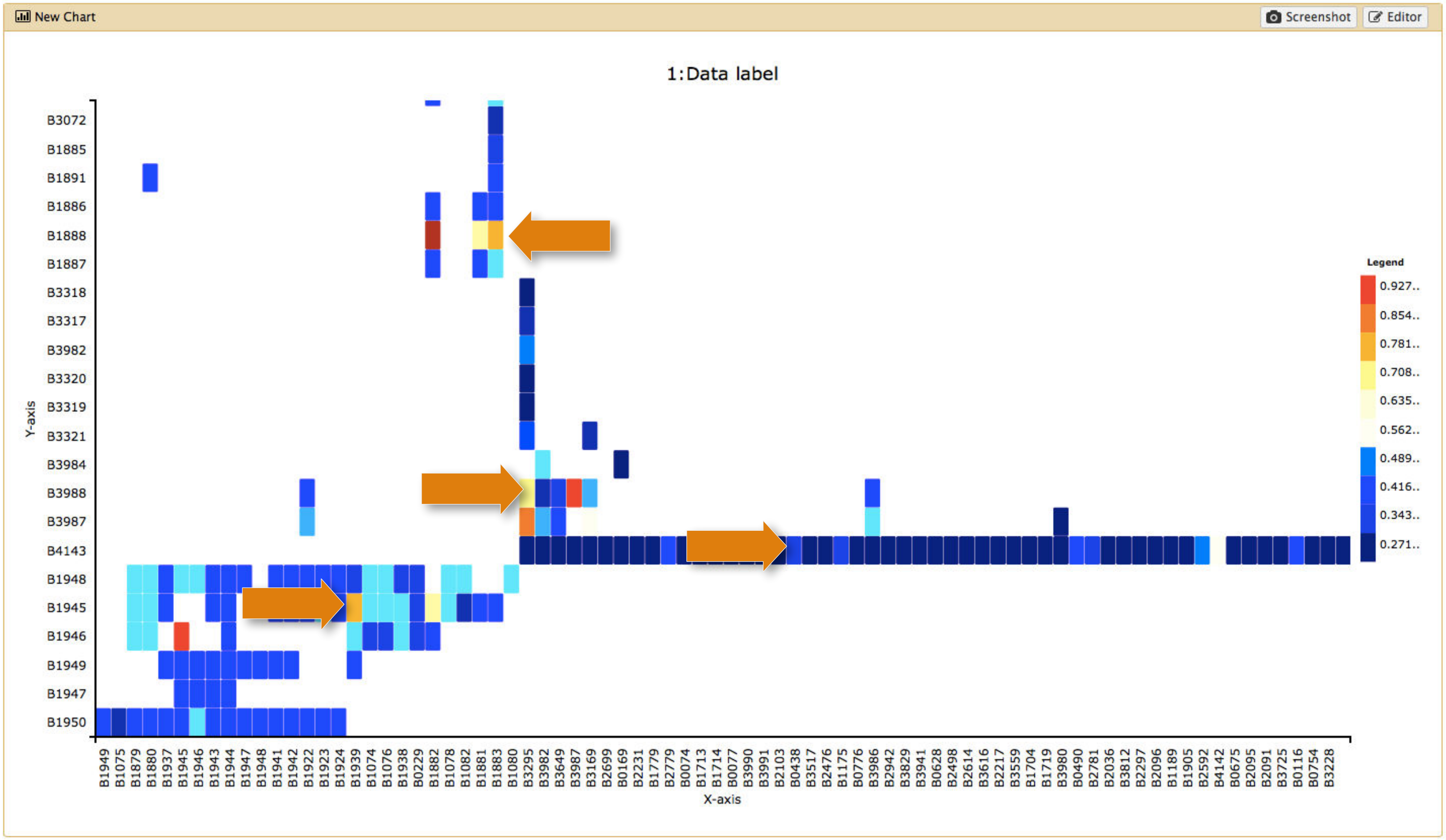
1. groEL - Stress factor RpoS regulon in exponential-phase bacteria
Annotation: groEL, molecular chaperone GroEL (multiple annotations exist)
Organism: Escherichia coli K-12
Reporter: GPL199, mopA_**b4143**_at (ID_REF), GDS3123, 1037522 (Gene ID), 913705 (Gene ID), 948665 (Gene ID), 959980 (Gene ID), **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

Database: Select

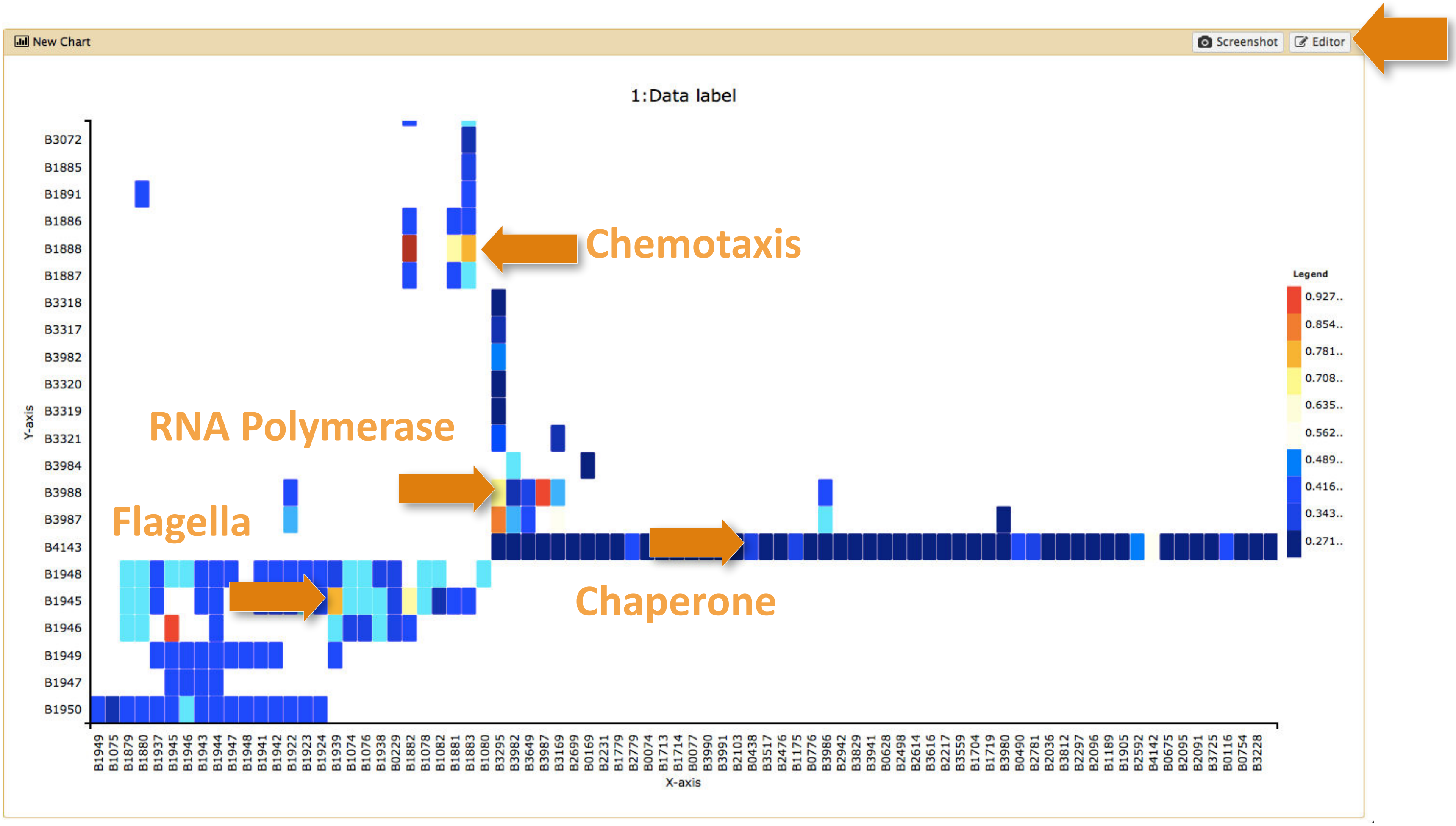
b4143[All Fields]

Cluster selection and analysis



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

Identified protein categories



Chemotaxis

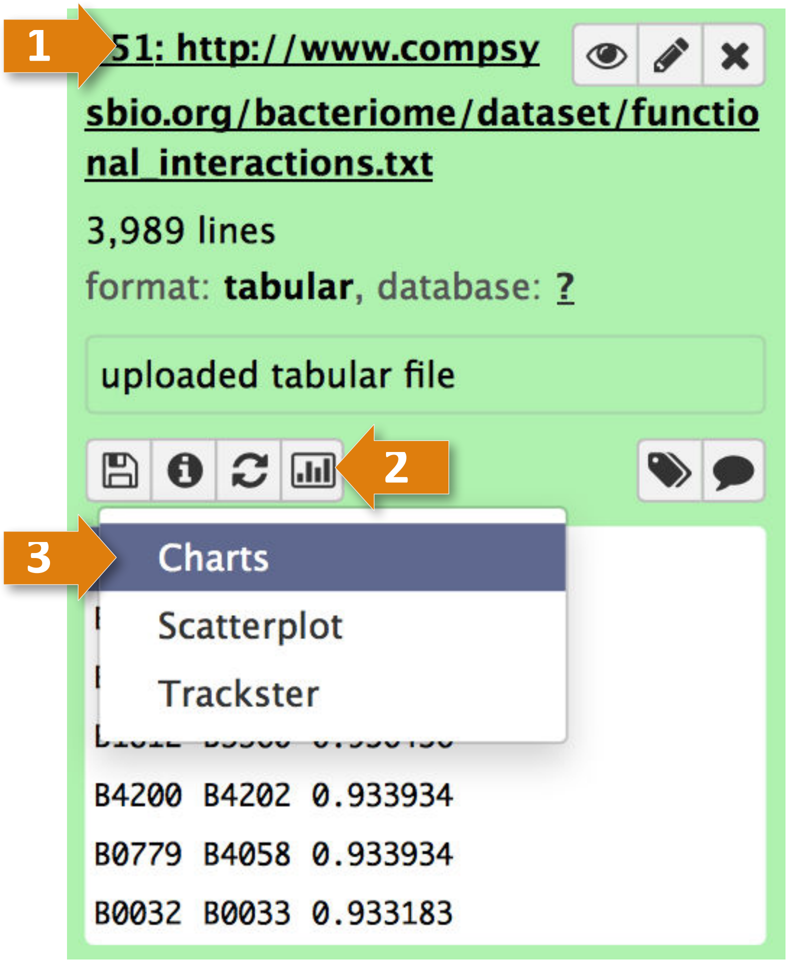
RNA Polymerase

Flagella

Chaperone

Please return to the **Editor**.

Make a new chart (3 of 4)



51: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

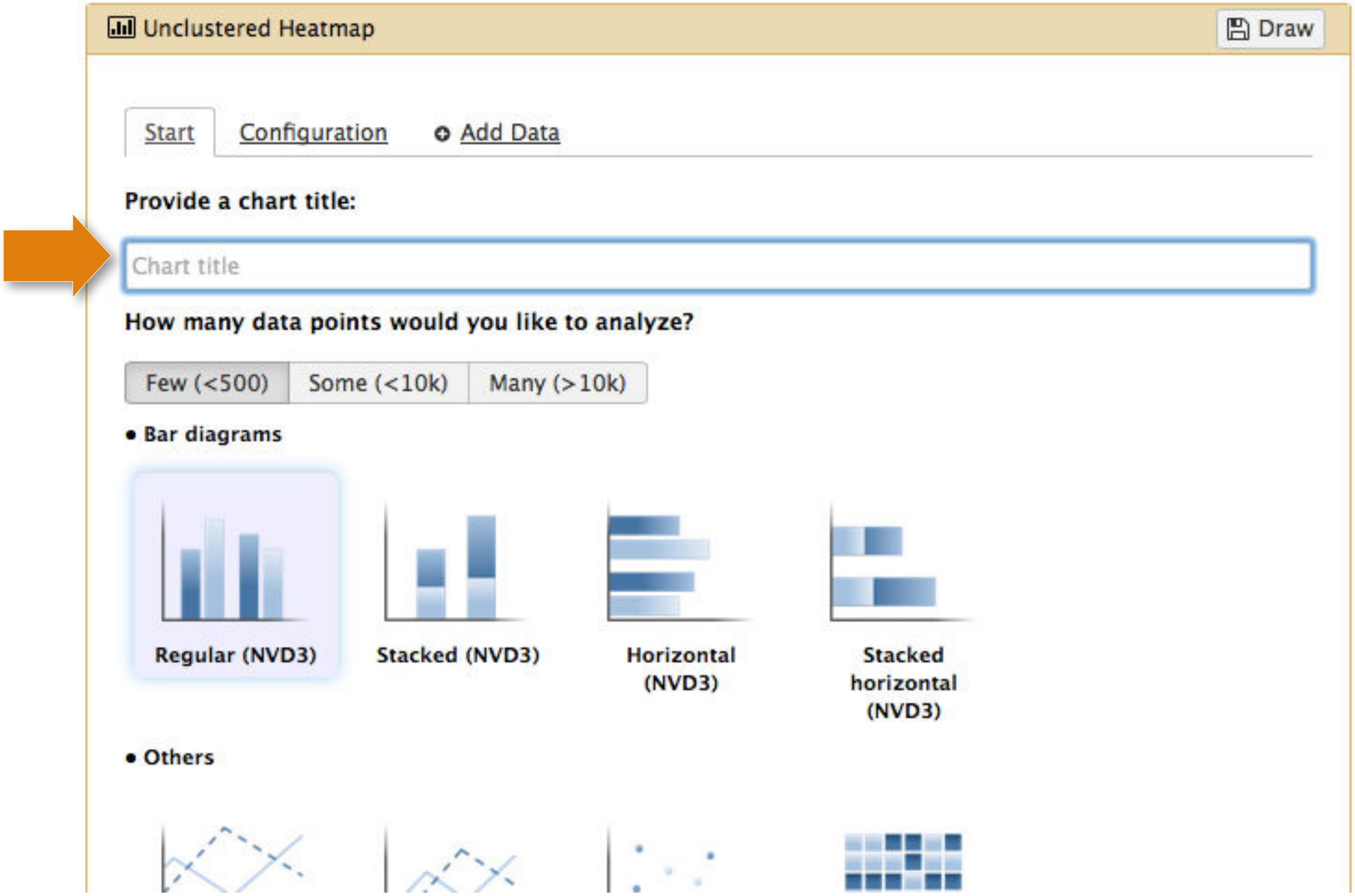
format: **tabular**, database: ?

uploaded tabular file

```
B4200  B4202  0.933934
B0779  B4058  0.933934
B0032  B0033  0.933183
```

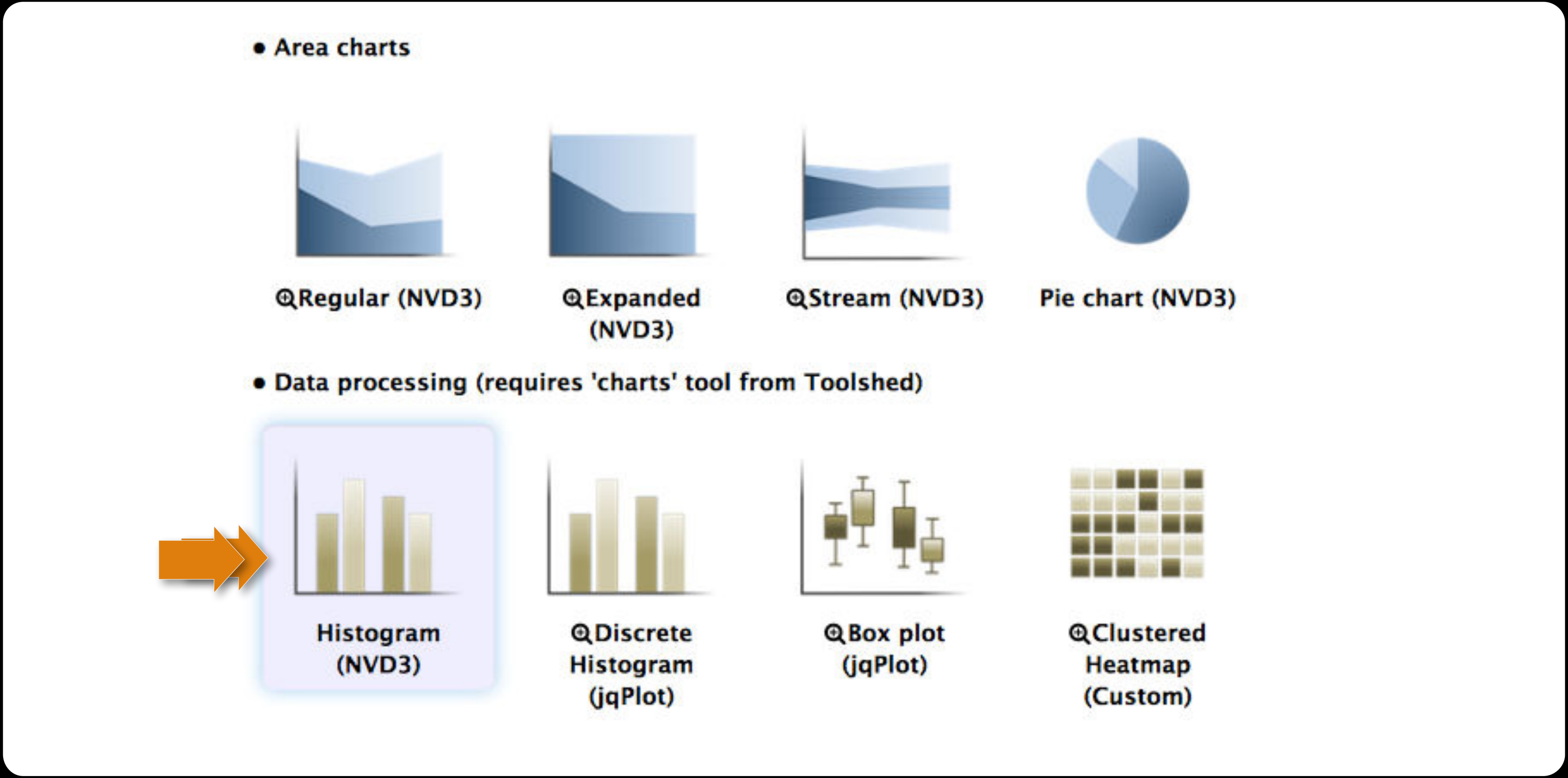
Select your **Dataset** and click on the **Visualization Icon** then select **Charts**.

Give your chart a name



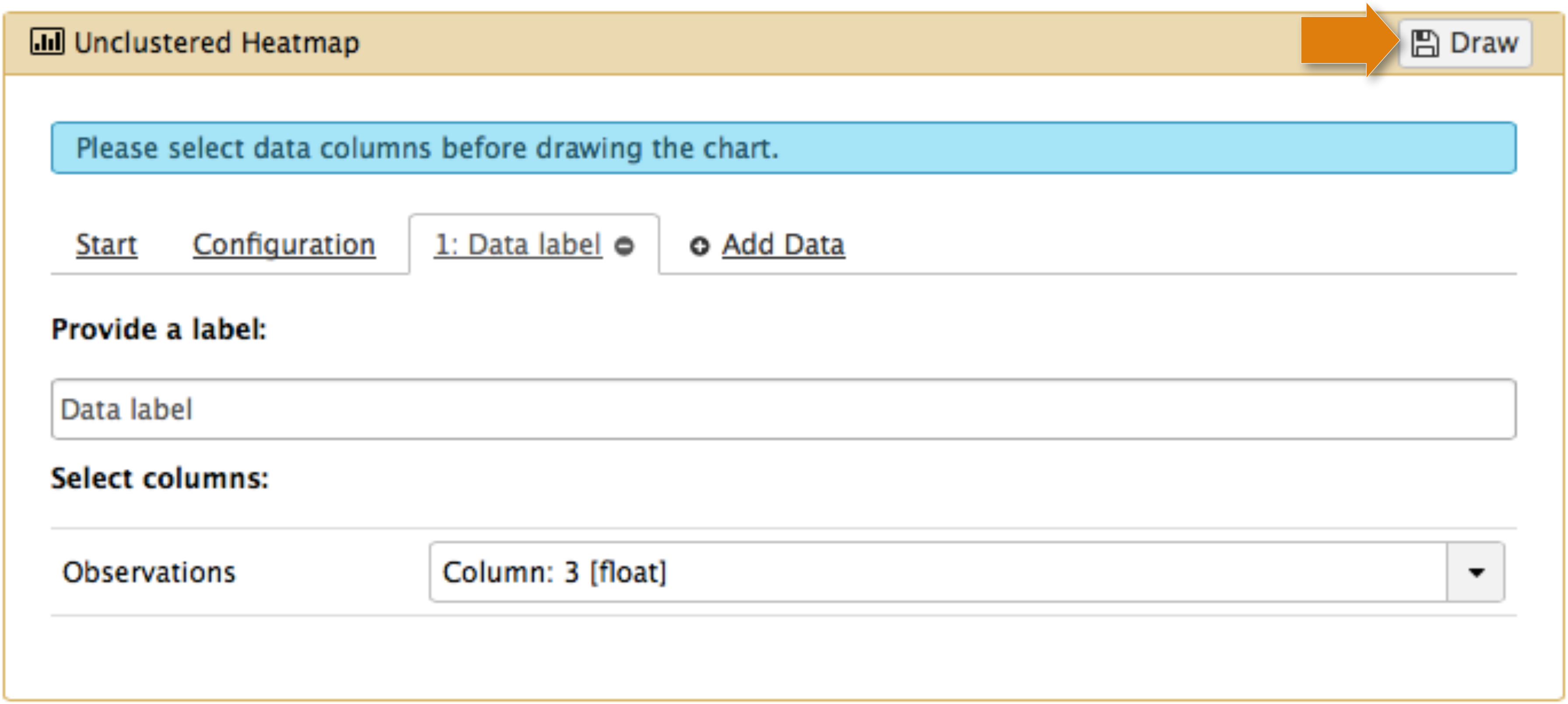
Name your chart **Score Histogram**.

Analyze the score distribution



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

Give your chart a name



Please select data columns before drawing the chart.

Provide a label:

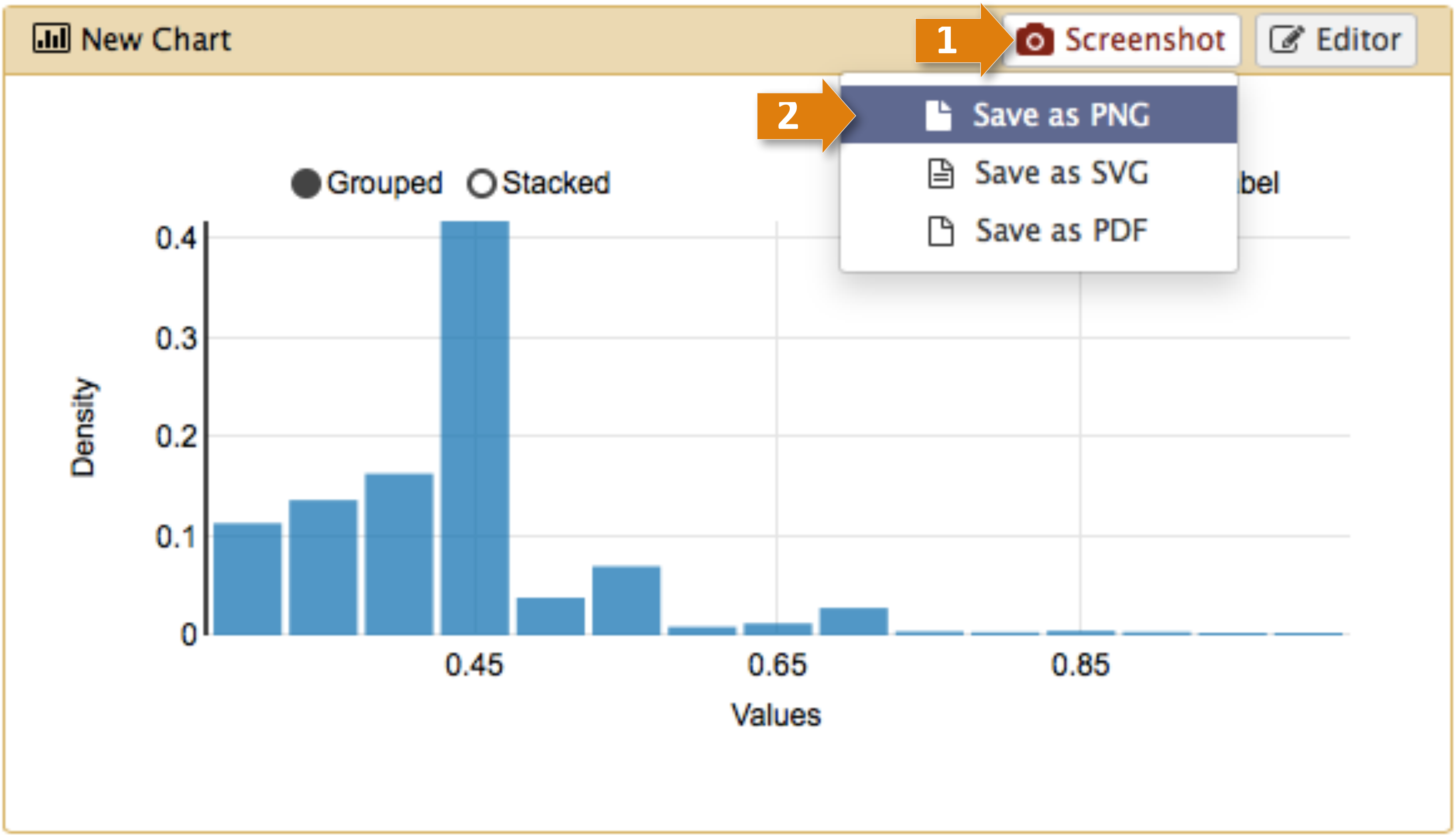
Data label

Select columns:

Observations: Column: 3 [float]

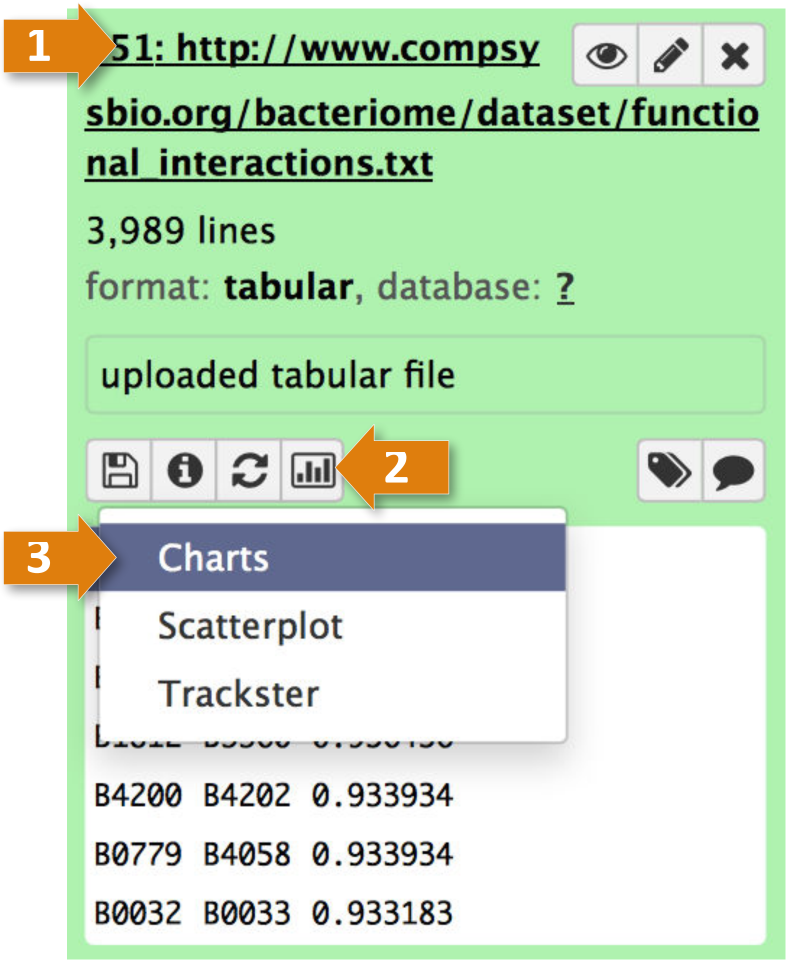
Click on **Draw**.

Export as PNG



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

Make a new chart (4 of 4)



51: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

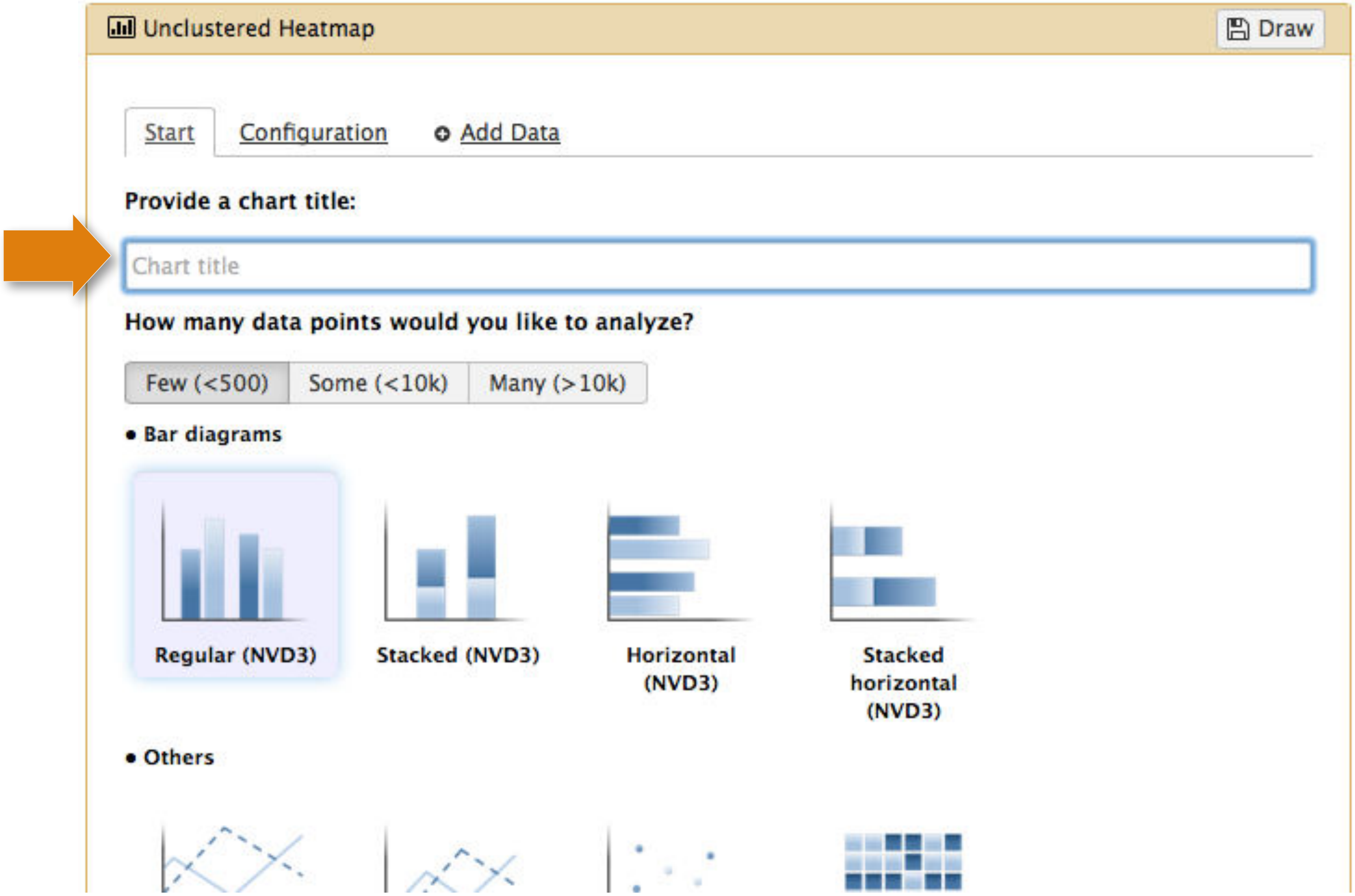
format: **tabular**, database: ?

uploaded tabular file

```
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



Provide a chart title:

Chart title

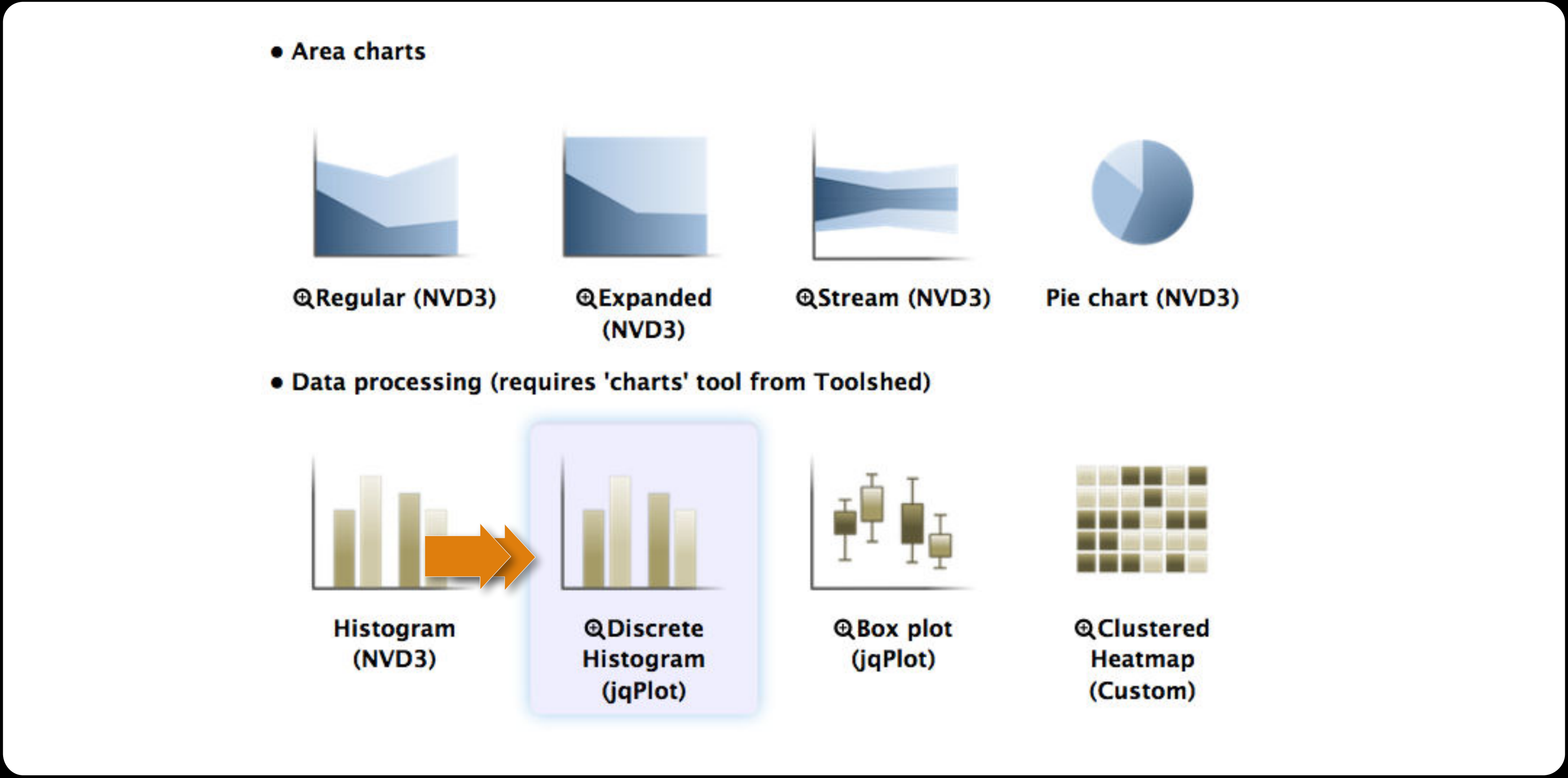
How many data points would you like to analyze?

- **Bar diagrams**

- **Others**

Name your chart **Discrete Histogram**.

Analyze the protein distribution

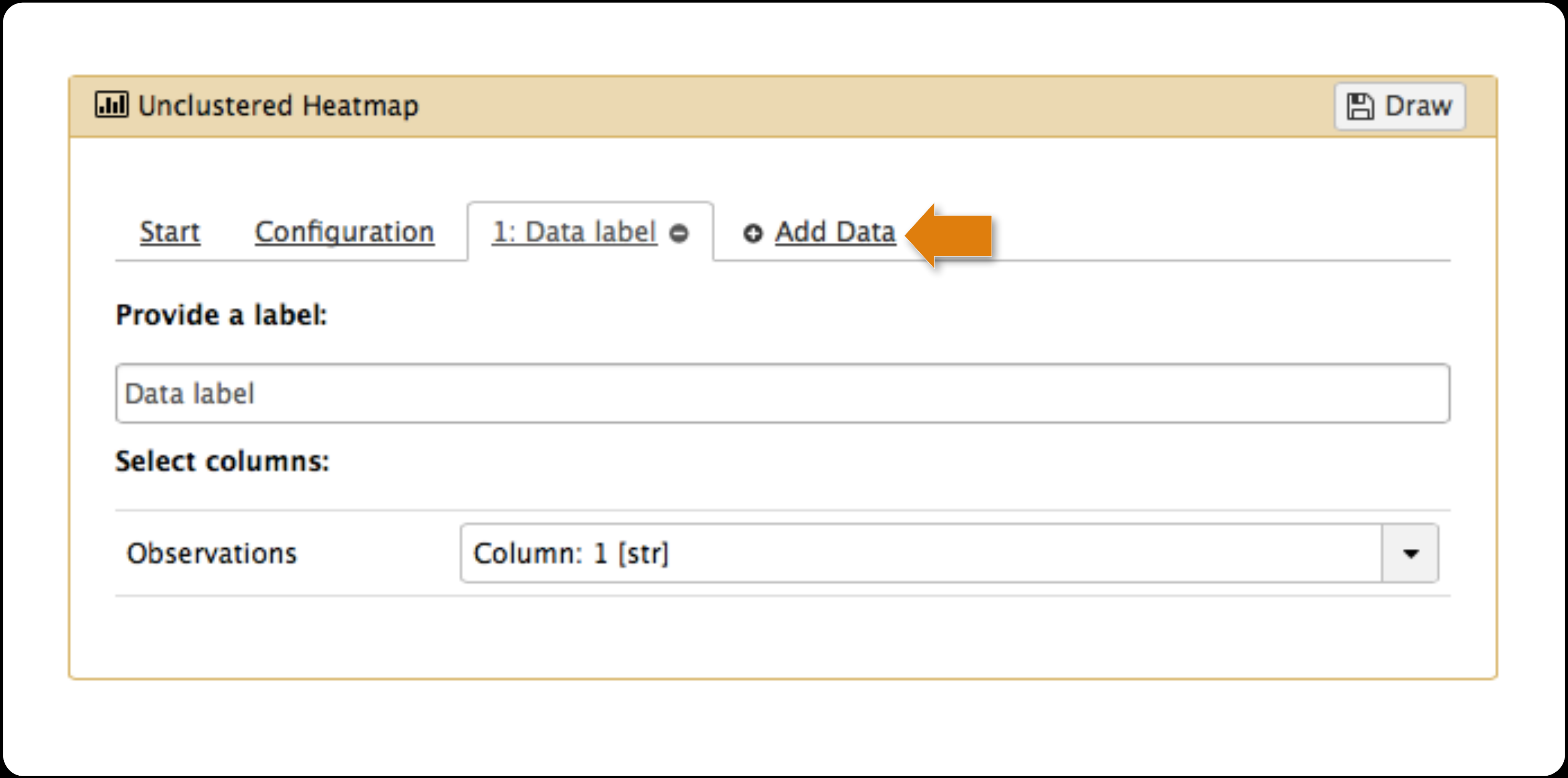


- Area charts

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

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

Add more data



Provide a label:

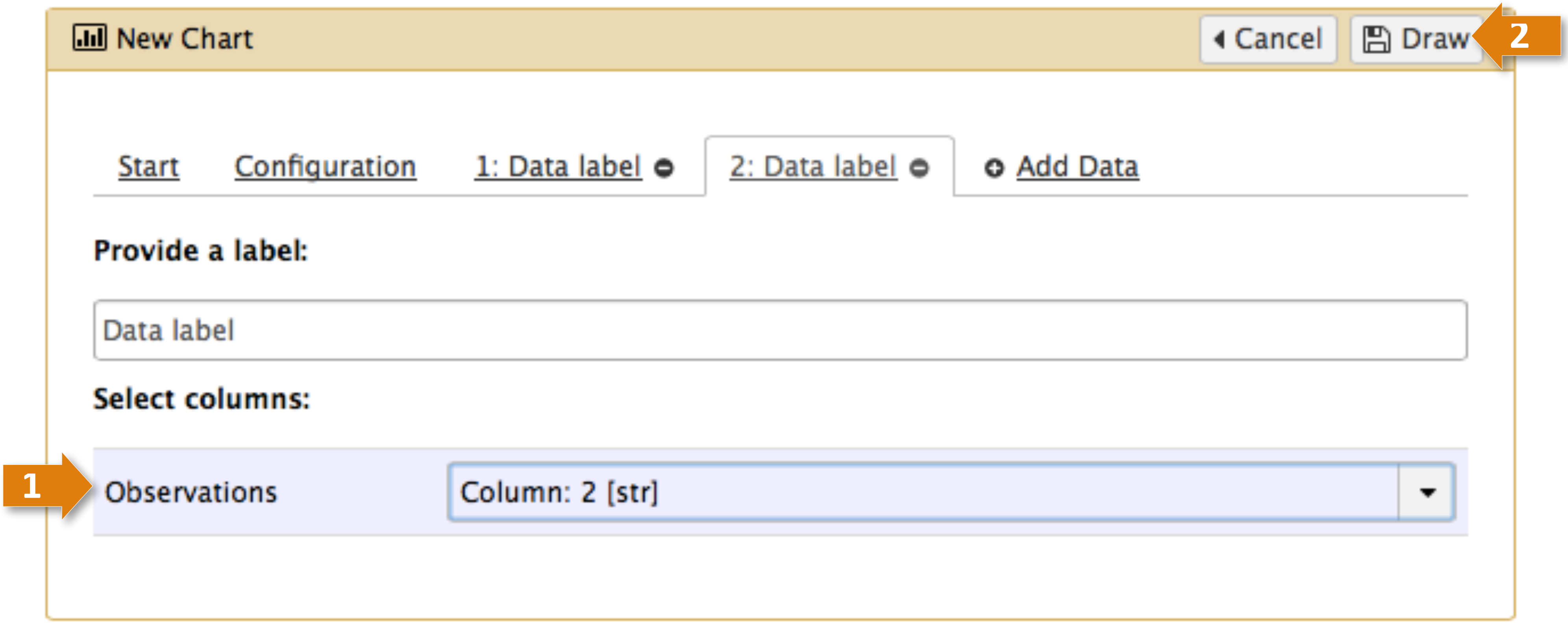
Data label

Select columns:

Observations: Column: 1 [str]

Click on **Add Data**.

Select a second data group



Provide a label:

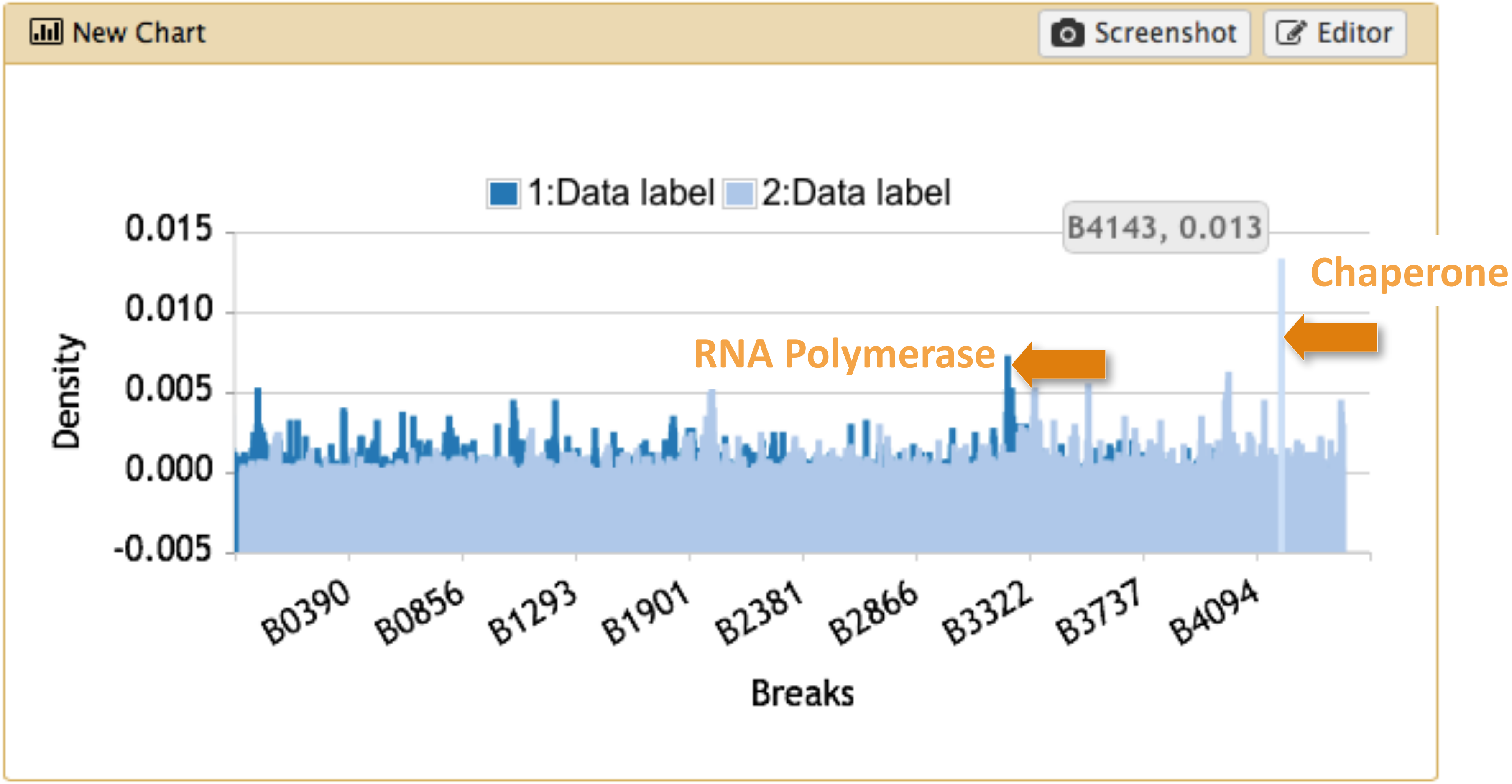
Data label

Select columns:

Observations: Column: 2 [str]

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

Which proteins have most interactions?



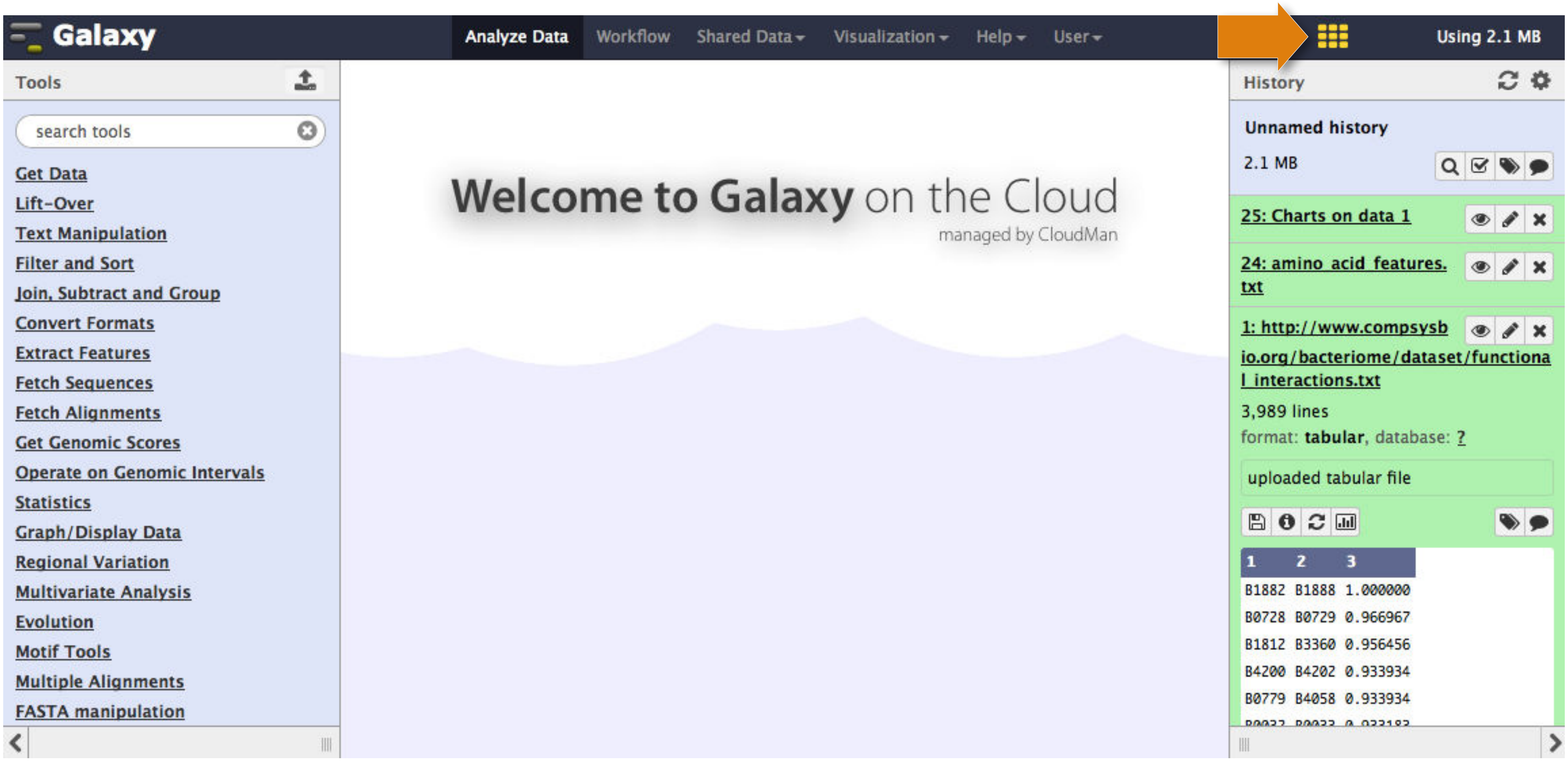
RNA Polymerase

Chaperone

Done with Part I.

Scratchbook

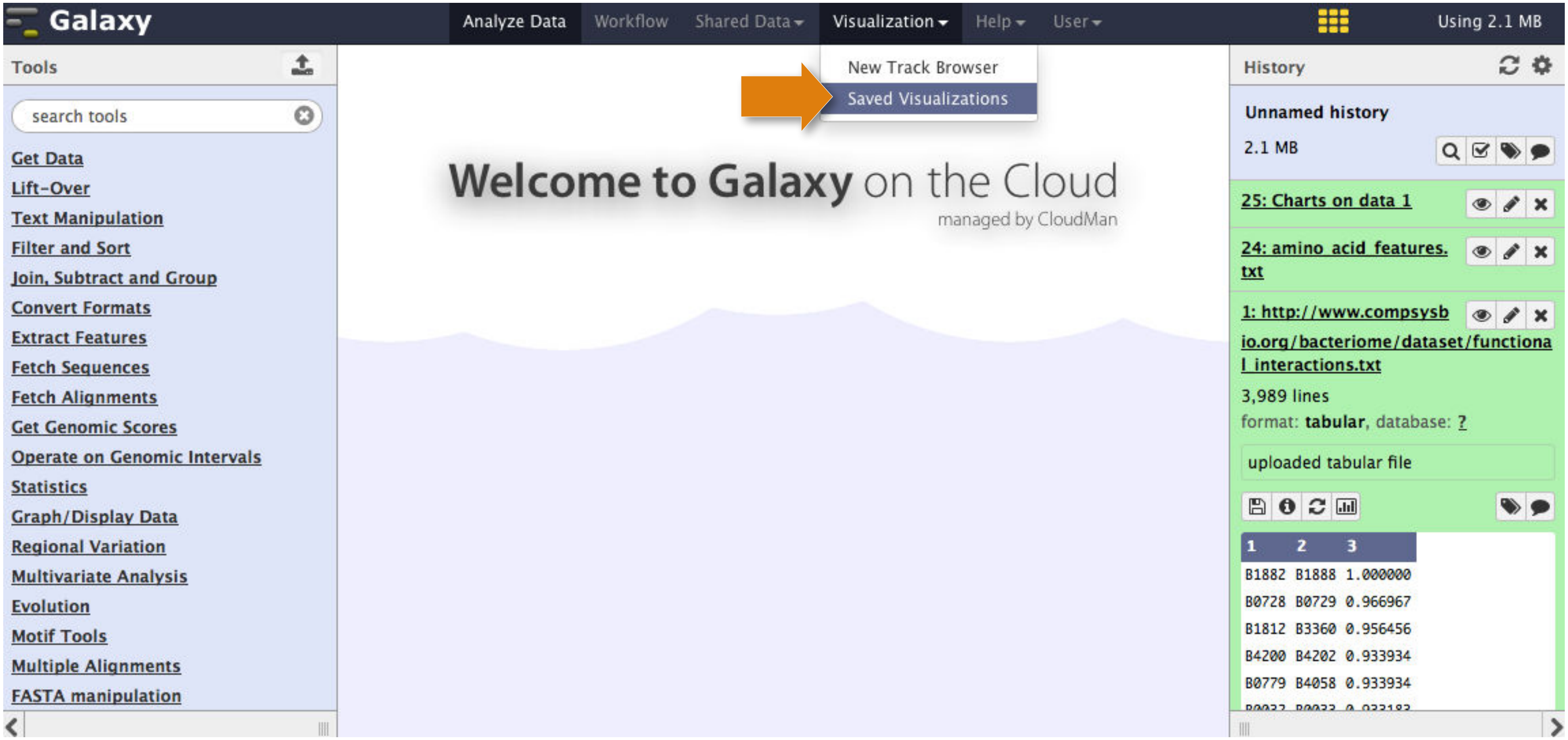
Activate the Scratchbook



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

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

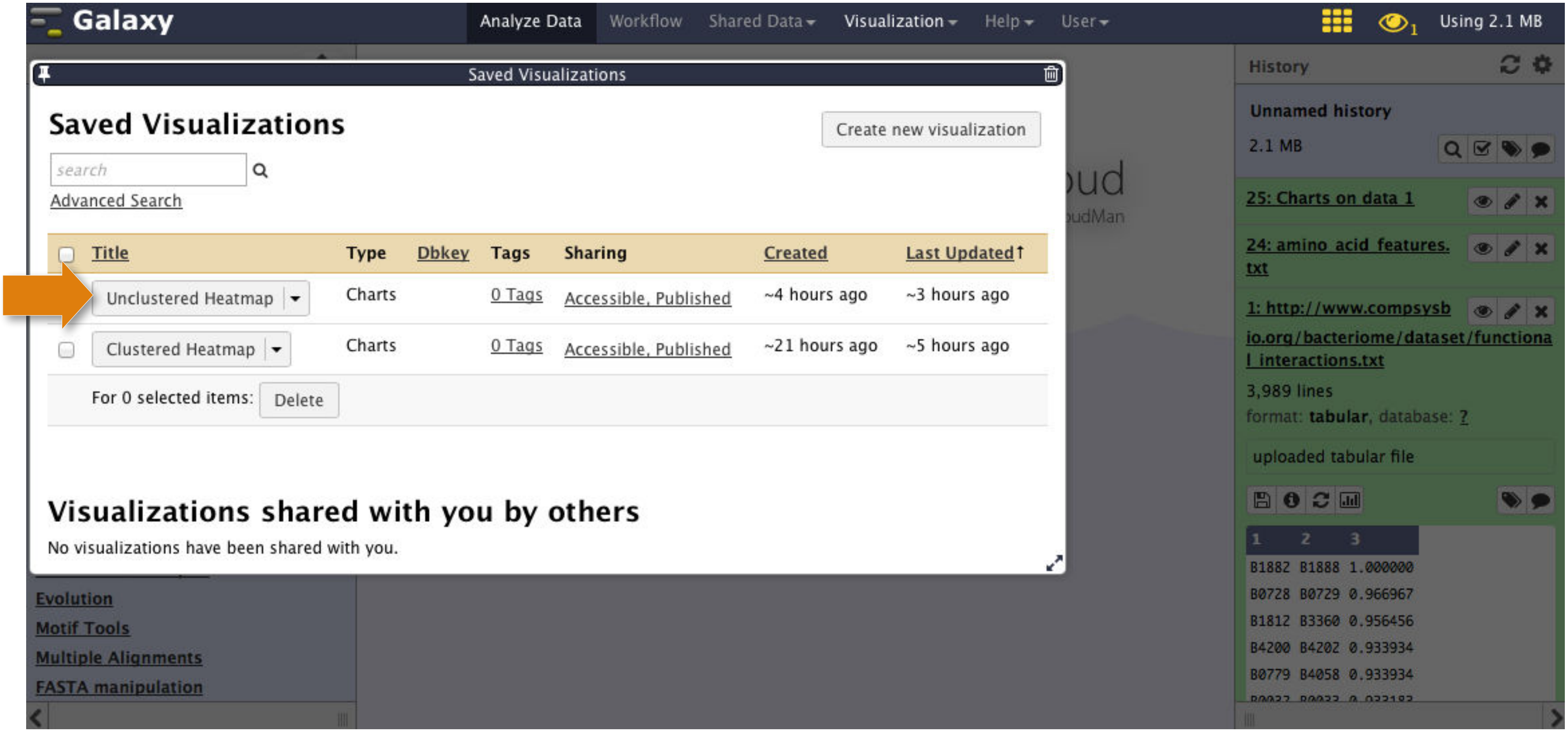
Activate the Scratchbook



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

Click on **Saved Visualizations**.

Activate the Scratchbook



Saved Visualizations

☐ 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: Delete

Visualizations shared with you by others

No visualizations have been shared with you.

Unnamed history

2.1 MB

25: Charts on data 1

24: amino acid features.txt

1: http://www.compsysbio.org/bacteriome/dataset/functional_interactions.txt

3,989 lines

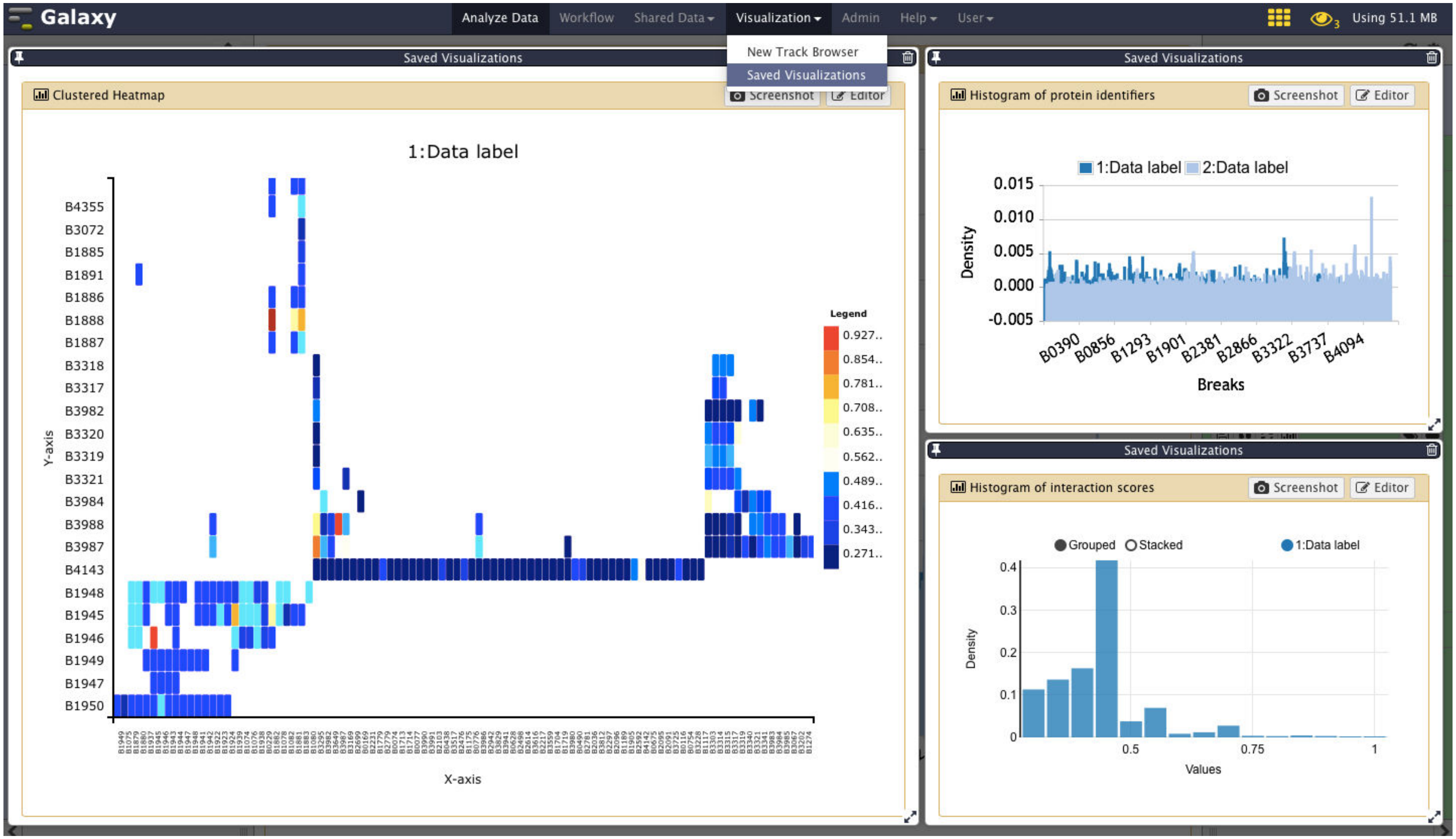
format: tabular, database: ?

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
```

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

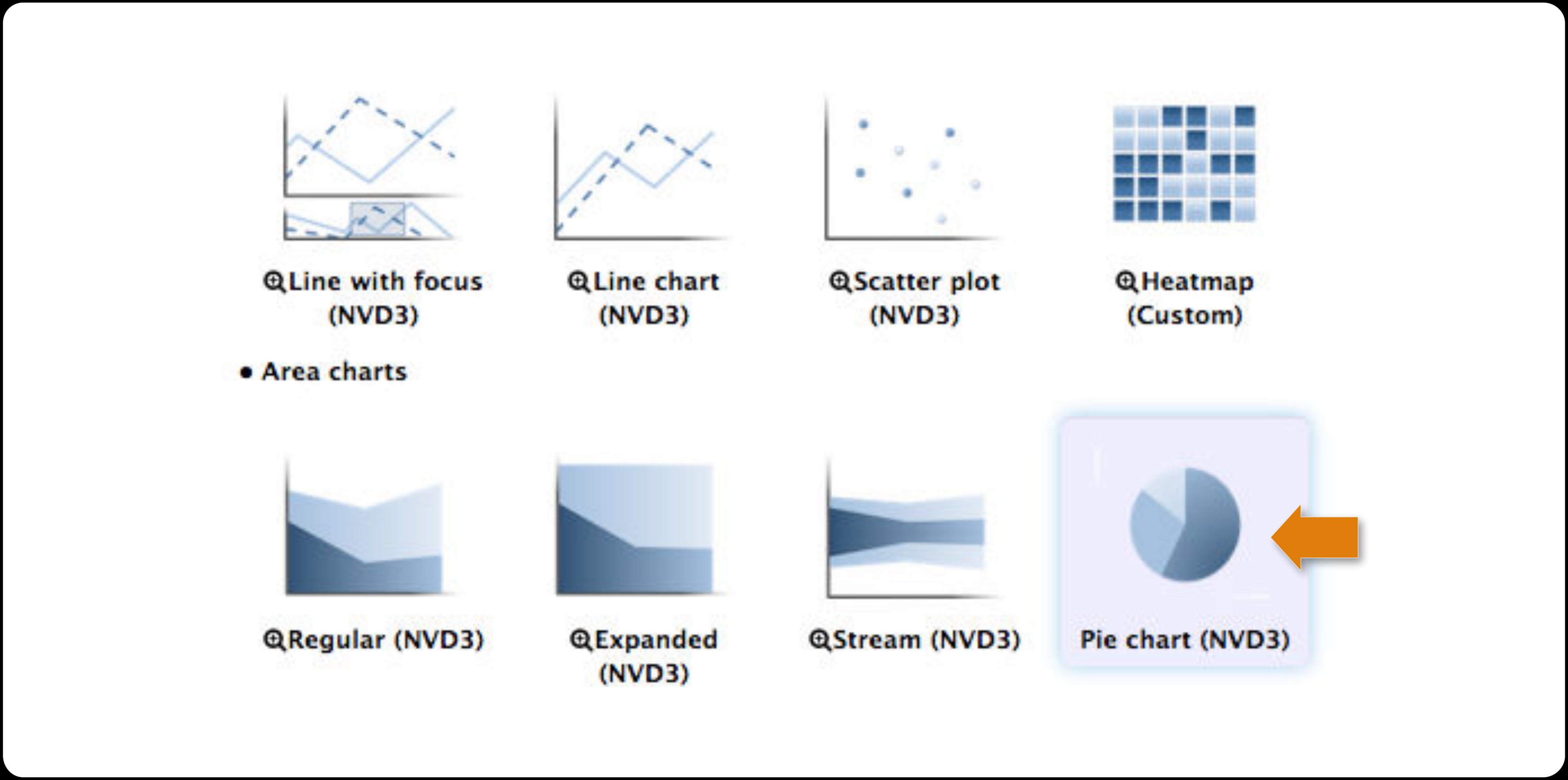
Scratchbook for multiple charts



Resize all visualizations so they fit into the screen.

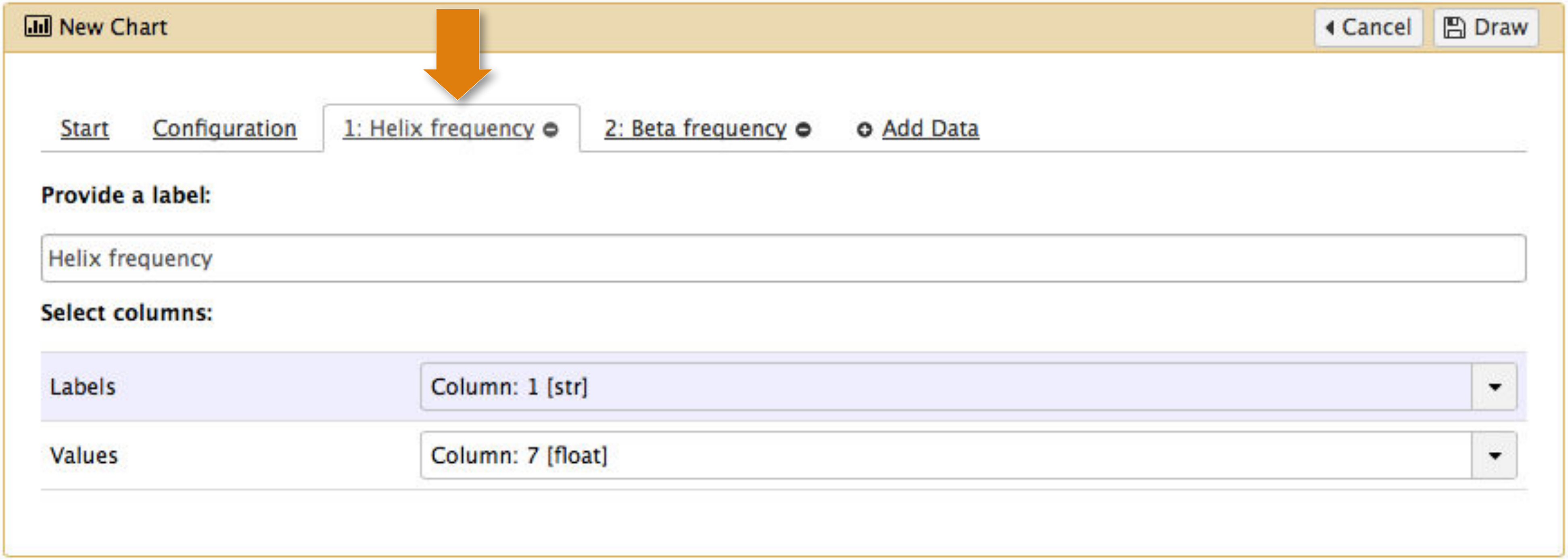
More Examples

Create a pie chart



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

Add first data group



Provide a label:

Helix frequency

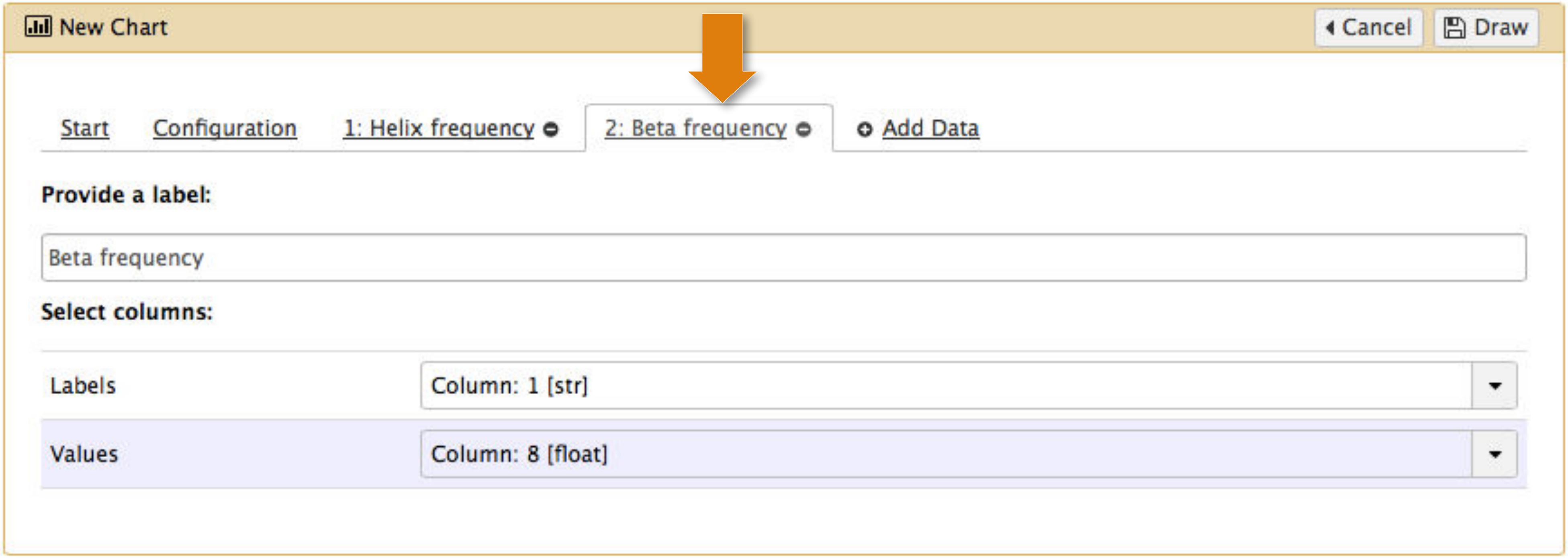
Select columns:

Labels: Column: 1 [str]

Values: Column: 7 [float]

Configure the **Helix frequency** column.

Add second data group



Provide a label:

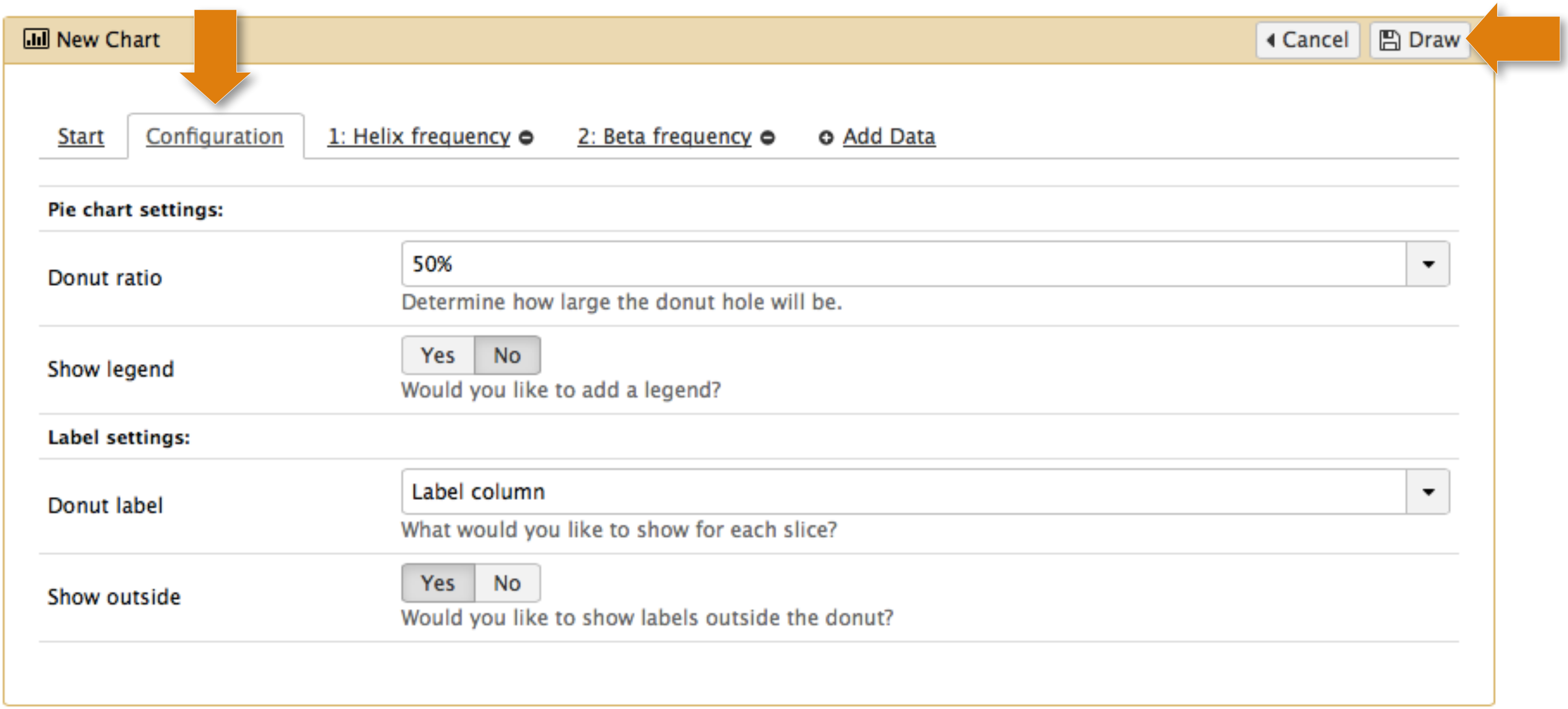
Beta frequency

Select columns:

Labels	Column: 1 [str]
Values	Column: 8 [float]

Configure the **Beta frequency** column.

Configure the pie chart



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?

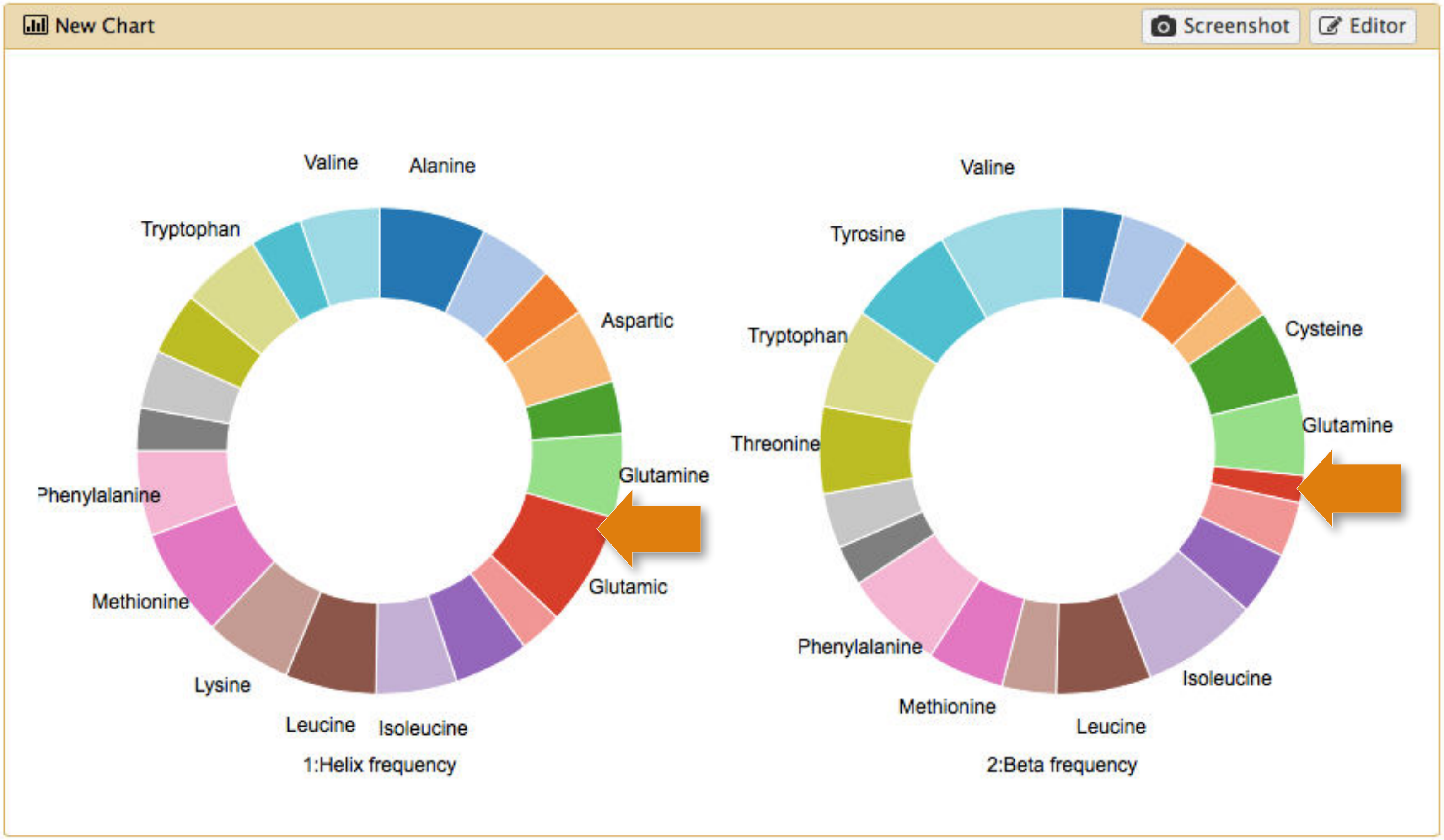
Label settings:

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?

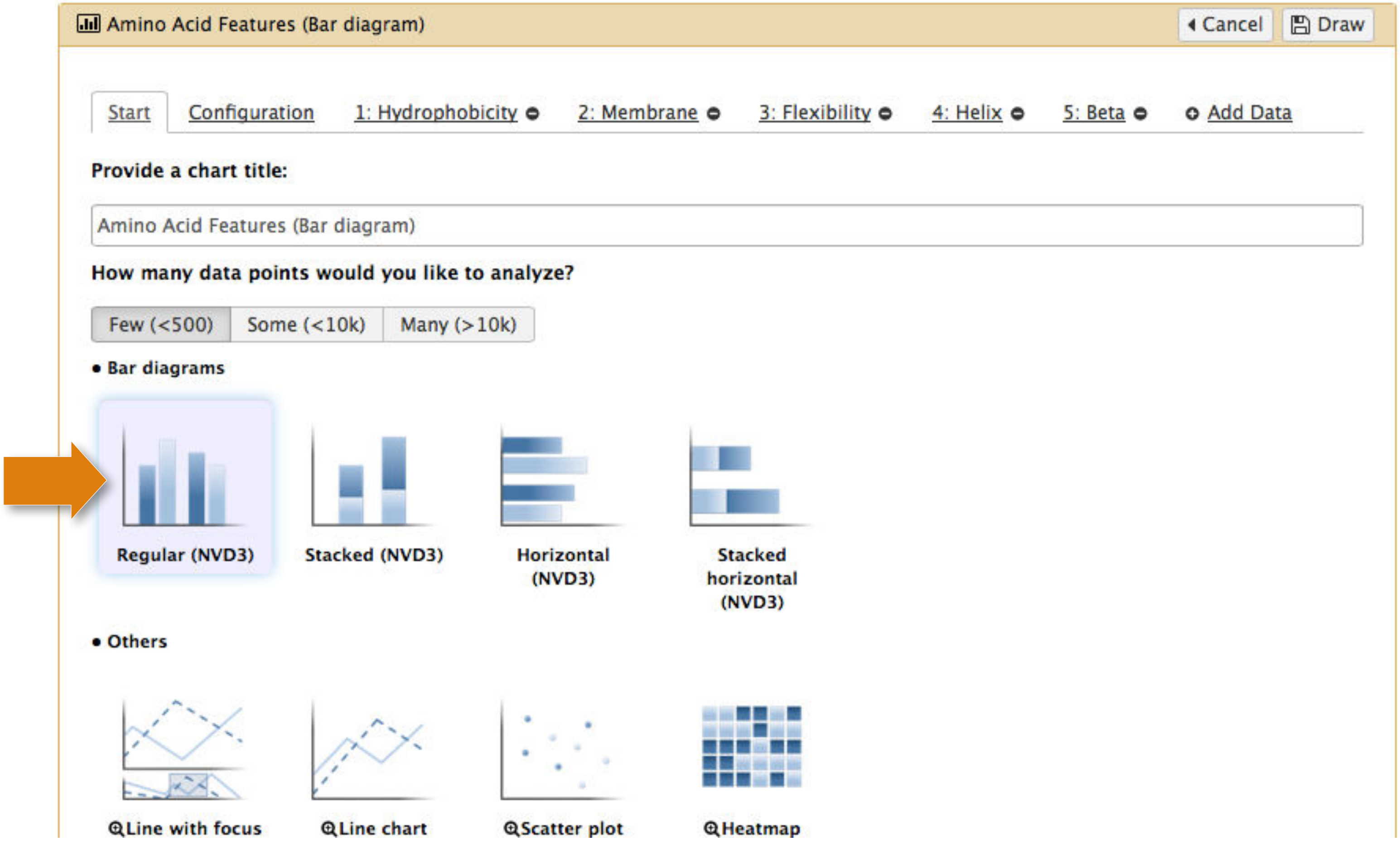
Configure the **Pie chart** as shown above. Then, click on **Draw**.

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



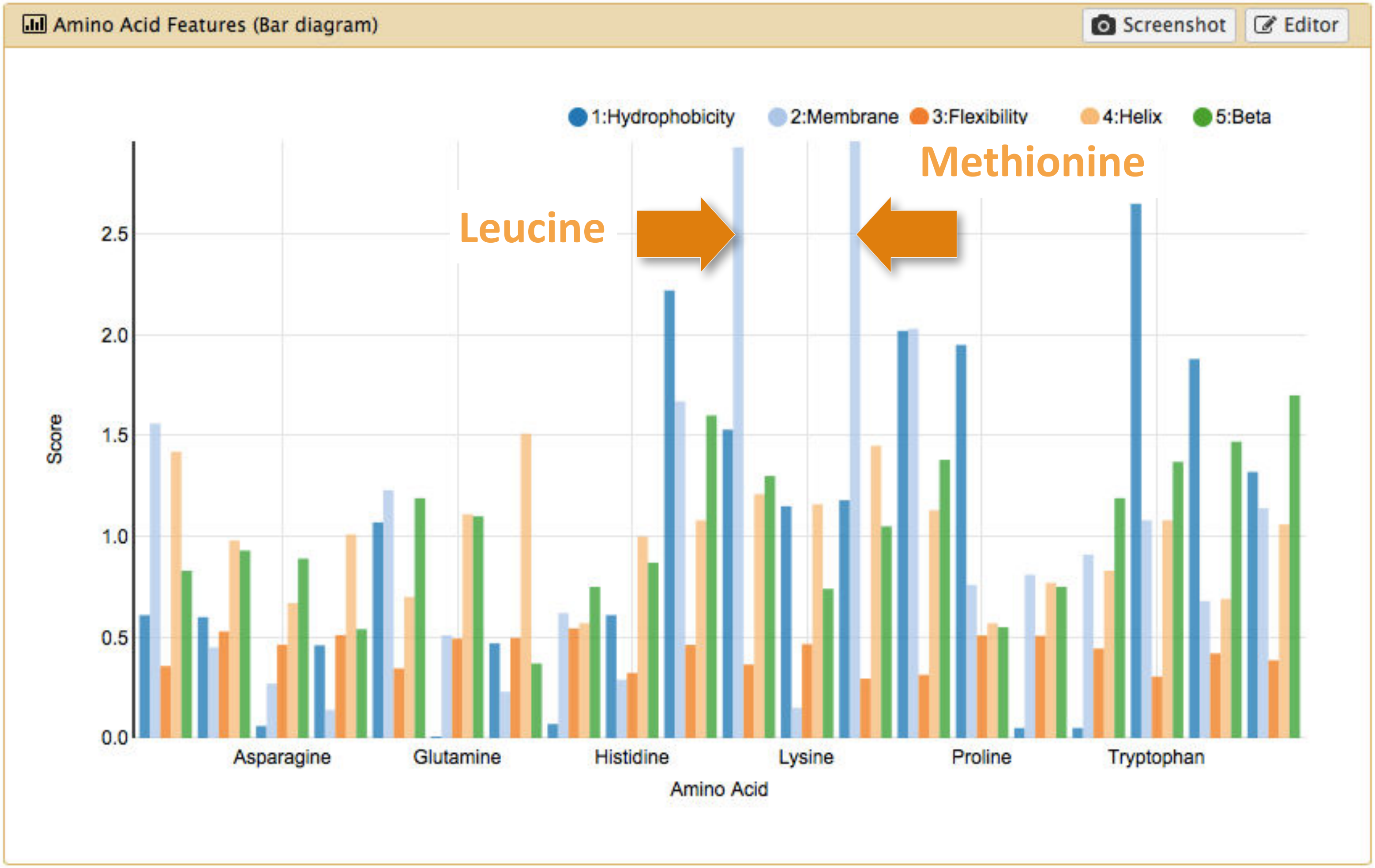
Amino Acid Features (Bar diagram)

Provide a chart title: Amino Acid Features (Bar diagram)

How many data points would you like to analyze? Few (<500)

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