Introduction to Galaxy

University of Utah
November 17, 2016

Dave Clements
Galaxy Team
Johns Hopkins University
http://galaxyproject.org/

THE UNIVERSITY OF UTAH®

DEPARTMENT OF
Biomedical Informatics

Galaxy PROJECT

#usegalaxy @galaxyproject

Slides URL

bit.ly/gxybmi6030slides

Goals

Provide an introduction to using Galaxy for bioinformatic analysis. Demonstrate how Galaxy can help you explore and learn options, perform analysis, and then share, repeat, and reproduce your analyses.

What is Galaxy?

Keith Bradnam's definition:

"A web-based platform that provides a simplified interface to many popular bioinformormatics tools."

From

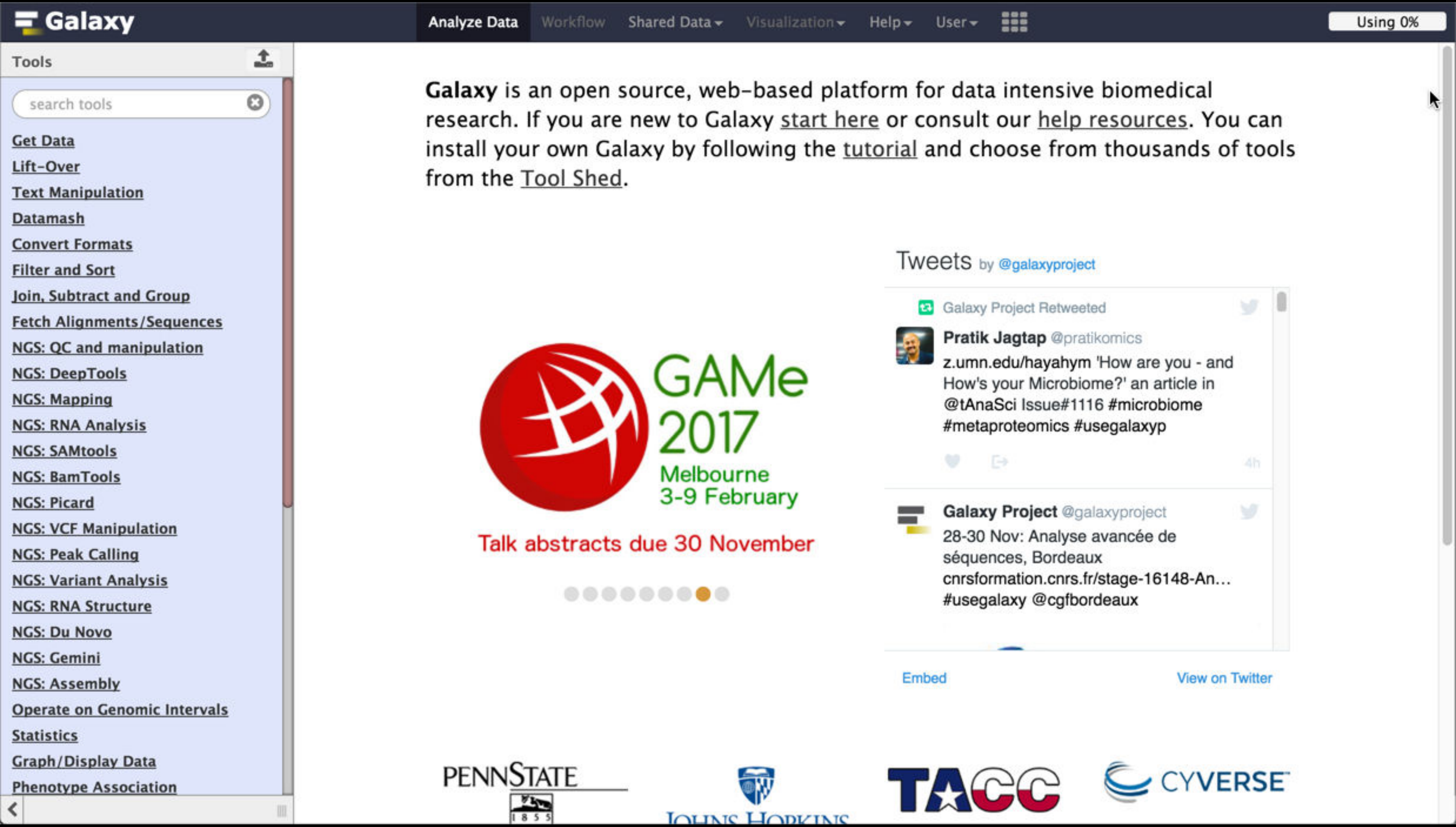
"13 Questions You May Have About Galaxy"

http://bit.ly/13questions

Galaxy is available several ways ...

http://galaxyproject.org

As a free for everyone service on the web: usegalaxy.org



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

Galaxy Project Retweeted

Pratik Jagtap @pratikomics

z.umn.edu/hayahym 'How are you - and How's your Microbiome?' an article in @tAnaSci Issue#1116 #microbiome #metaproteomics #usegalaxyp

Galaxy Project @galaxyproject

28-30 Nov: Analyse avancée de séquences, Bordeaux cnrsformation.cnrs.fr/stage-16148-An… #usegalaxy @cgfbordeaux

Galaxy is available as Open Source Software

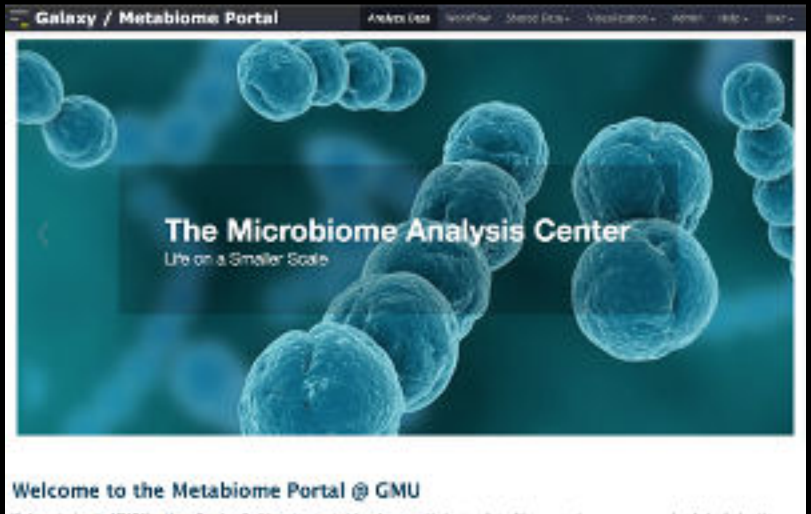
Galaxy is installed in locations around the world.

http://getgalaxy.org

Explore the Galaxy with

RNA-Rocket

PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL



The Microbiome Analysis Center
Life on a Smaller Scale

Welcome to the Metabiome Portal @ GMU

香港中文大學 - 華大基因跨組學創新研究院
CUHK-BGI Innovation Institute of Trans-Omics
华大基因 BGI

(GIGA)^n Galaxy
by CBIIT

Integrated publishing of workflows from (GIGA)^n SCIENCE

Cistrome

A Galaxy Server dedicated to ChIP-* analysis

080+

Public Galaxy Servers
and *still* counting

The Genomic HyperBrowser

Powered by Galaxy

SCDE
STEM CELL DISCOVERY ENGINE

Experiments Connected

Whale Shark Galaxy!

South Green
bioinformatics platform

Genomic analysis tools for southern and Mediterranean plants

bit.ly/gxyServers

Galaxy is available on the Cloud

openstack™
CLOUD SOFTWARE

amazon web services™

OpenNebula.org
The Open Source Toolkit for Cloud Computing

globus genomics

We are using this today

http://aws.amazon.com/education
http://globus.org/
http://wiki.galaxyproject.org/Cloud

Galaxy on the Cloud: Galaxy CloudMan

http://usegalaxy.org/cloud

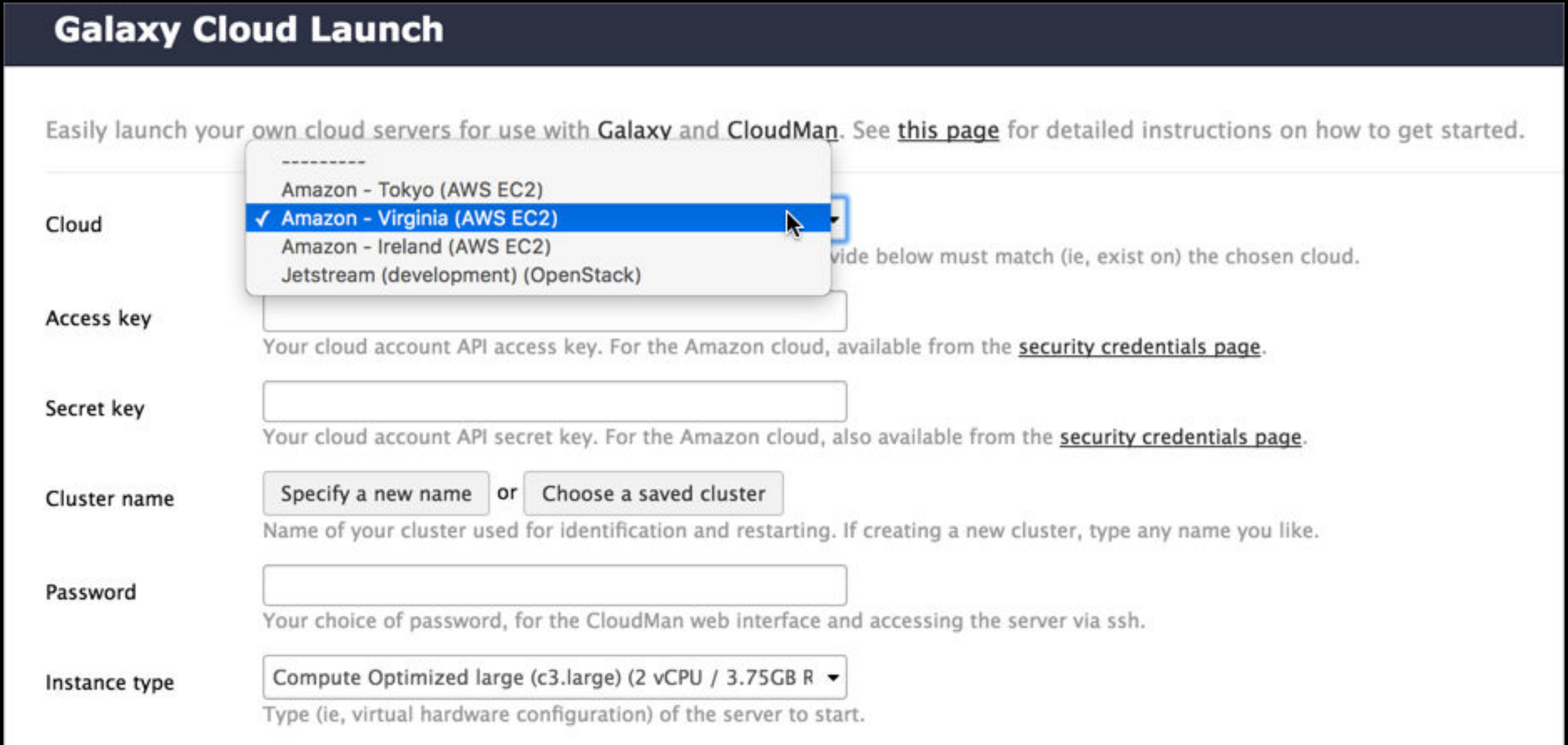
- Start with a **fully configured and populated** (tools and data) Galaxy instance.
- Allows you to scale up and down your compute assets as needed.
- Someone else manages the data center

CLOUDMAN

Galaxy on the Cloud: CloudLaunch

https://launch.usegalaxy.org/

- Directly launch a Galaxy instance on AWS or Jetstream
- Uses CloudMan



Galaxy Cloud Launch

Easily launch your own cloud servers for use with Galaxy and CloudMan. See this page for detailed instructions on how to get started.

Cloud: Amazon - Virginia (AWS EC2)

…vide below must match (ie, exist on) the chosen cloud.

Access key

Your cloud account API access key. For the Amazon cloud, available from the security credentials page.

Secret key

Your cloud account API secret key. For the Amazon cloud, also available from the security credentials page.

Cluster name

Name of your cluster used for identification and restarting. If creating a new cluster, type any name you like.

Password

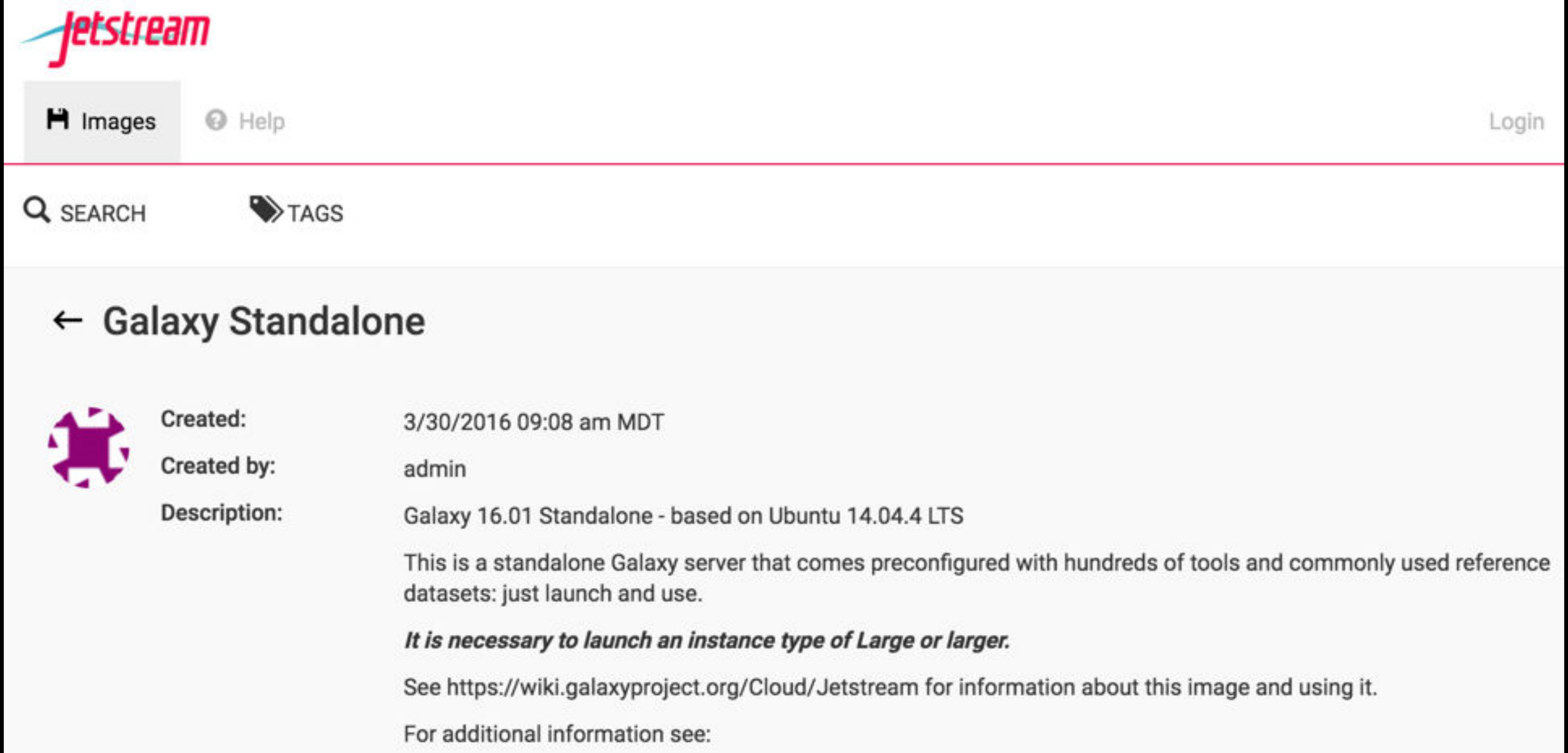
Your choice of password, for the CloudMan web interface and accessing the server via ssh.

Instance type: Compute Optimized large (c3.large) (2 vCPU / 3.75GB R

Type (ie, virtual hardware configuration) of the server to start.

Galaxy on the Cloud: Jetstream

https://wiki.galaxyproject.org/Cloud/Jetstream



← Galaxy Standalone

Created: 3/30/2016 09:08 am MDT

Created by: admin

Description: Galaxy 16.01 Standalone - based on Ubuntu 14.04.4 LTS

This is a standalone Galaxy server that comes preconfigured with hundreds of tools and commonly used reference datasets: just launch and use.

It is necessary to launch an instance type of Large or larger.

See https://wiki.galaxyproject.org/Cloud/Jetstream for information about this image and using it.

For additional information see:

US based researchers can request an XSEDE allocation and then run Galaxy on Jetstream

U XSEDE Champion: Anita Orendt

Basic Analysis

Which exons have most overlapping Repeats?

Use Human, HG38, GENCODE v24, Chromosome 22

bit.ly/gxybmi6030

Exons & Repeats: A General Plan

- Get some data
  - Get Data → UCSC Table Browser
- Identify which exons have Repeats
- Count Repeats per exon
- Visualize, save, download, ... exons with most Repeats

(~ http://usegalaxy.org/galaxy101)

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersecti
DNA sequence covered by a track. For help in using this application see Using the Table Browse
this form, the User's Guide for general information and sample queries, and the OpenHelix Table
presentation of the software features and usage. For more complex queries, you may want to us
To examine the biological function of your set through annotation enrichments, send the data to
for use with diverse computational tools. Refer to the Credits page for the list of contributors and
these data. All tables can be downloaded in their entirety from the Sequence and Annotation Do

clade: Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)

group: Genes and Gene Predictions **track:** GENCODE v24

table: knownGene

region: ○ genome ● positio chr22

identifiers (names/accessions):

filter:

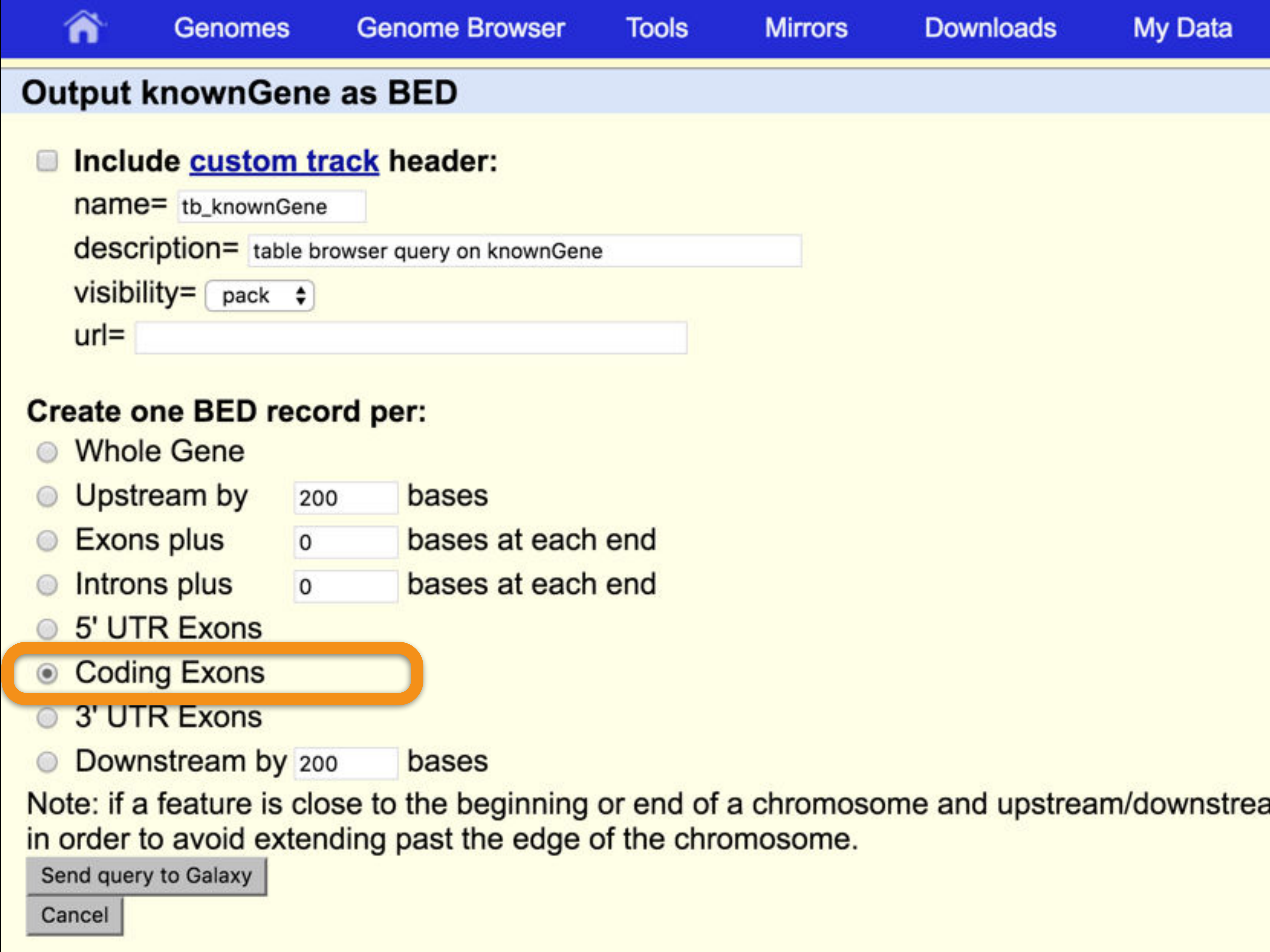
intersection:

correlation:

output format: BED - browser extensible data Send output to ☑ Galaxy ☐ GREAT ☐

output file: (leave blank to keep output in browser)

file type returned: ● plain text ○ gzip compressed



Output knownGene as BED

☐ **Include custom track header:**

name= tb_knownGene

description= table browser query on knownGene

visibility= pack

url=

Create one BED record per:

○ Whole Gene

○ Upstream by 200 bases

○ Exons plus 0 bases at each end

○ Introns plus 0 bases at each end

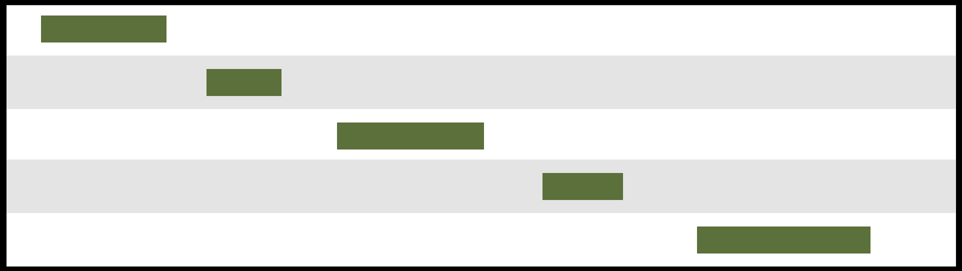
○ 5' UTR Exons

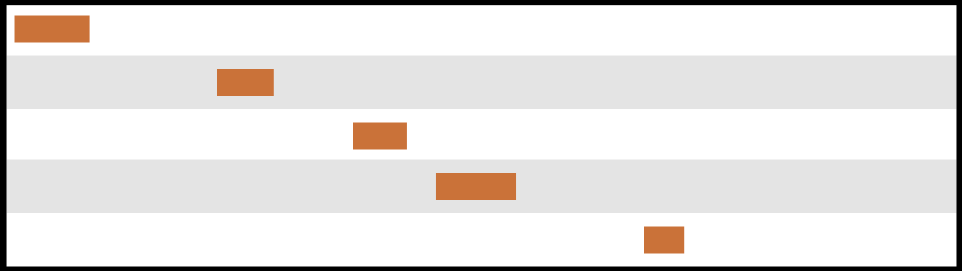
● Coding Exons

○ 3' UTR Exons

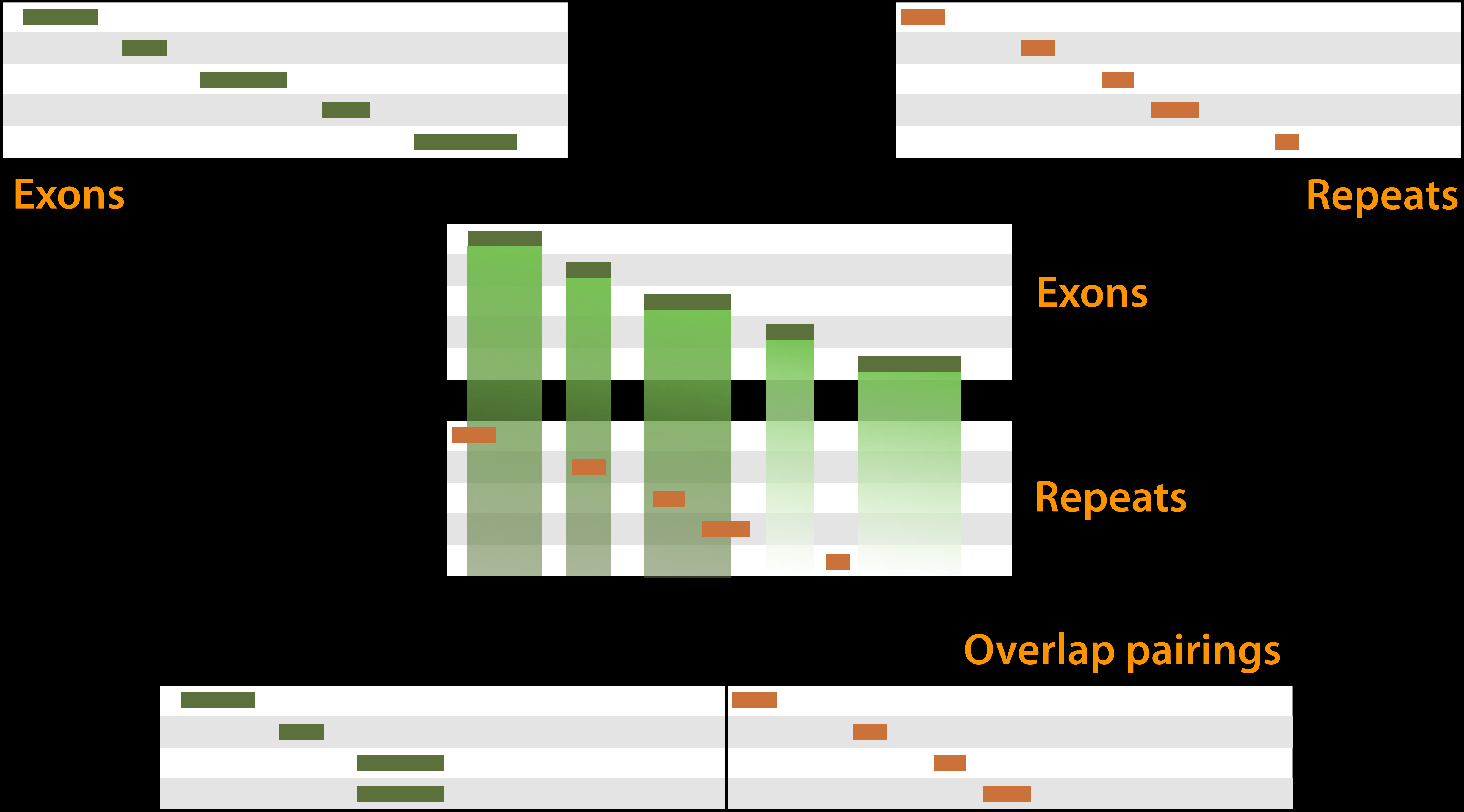
○ Downstream by 200 bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstrea
in order to avoid extending past the edge of the chromosome.


Exons

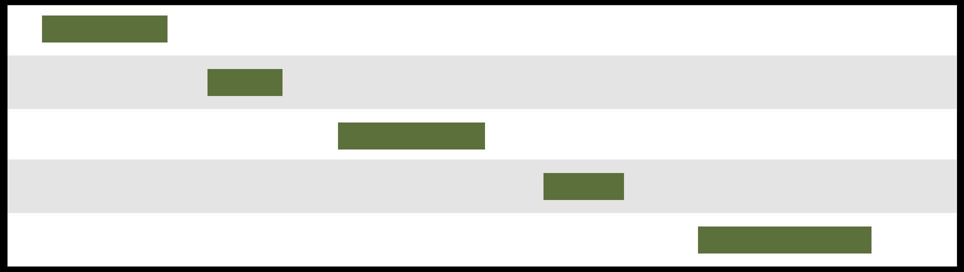

Repeats

(Identify which exons have Repeats)

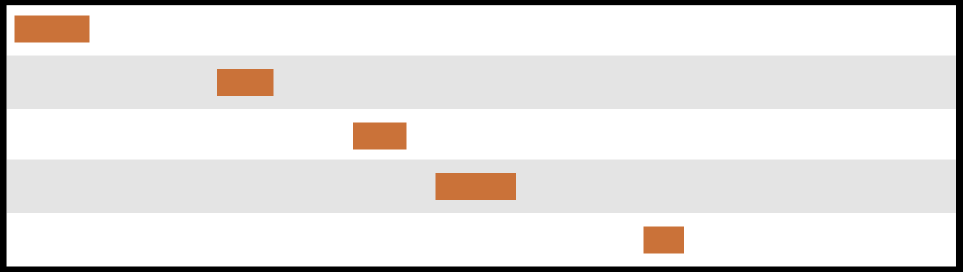


Operate on Genomic Intervals → **Join**

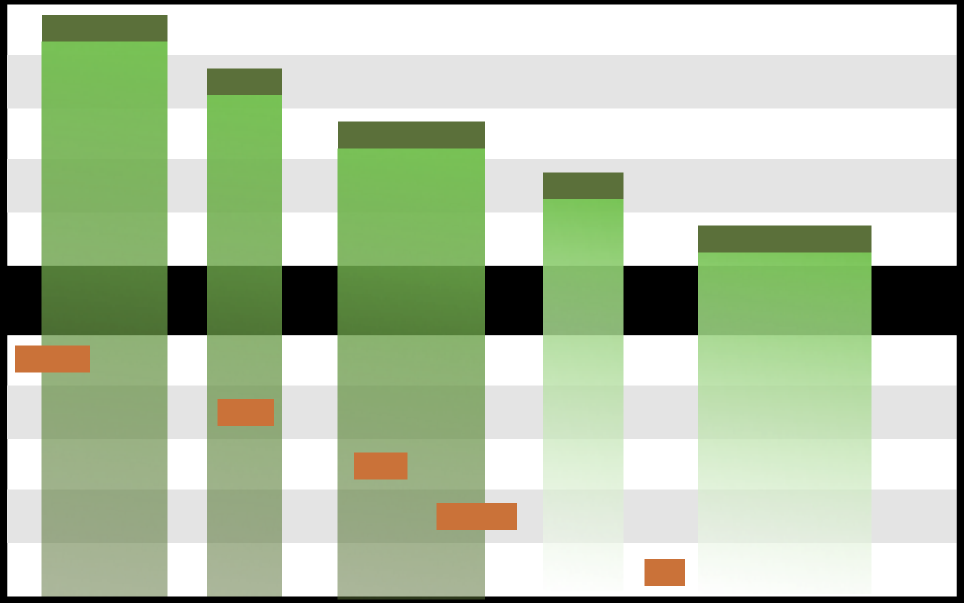
(Identify which exons have Repeats)



Exons



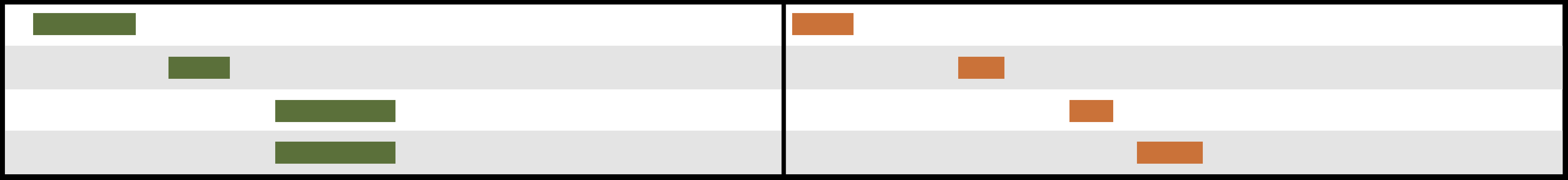
Repeats



Exons

Repeats

Overlap pairings



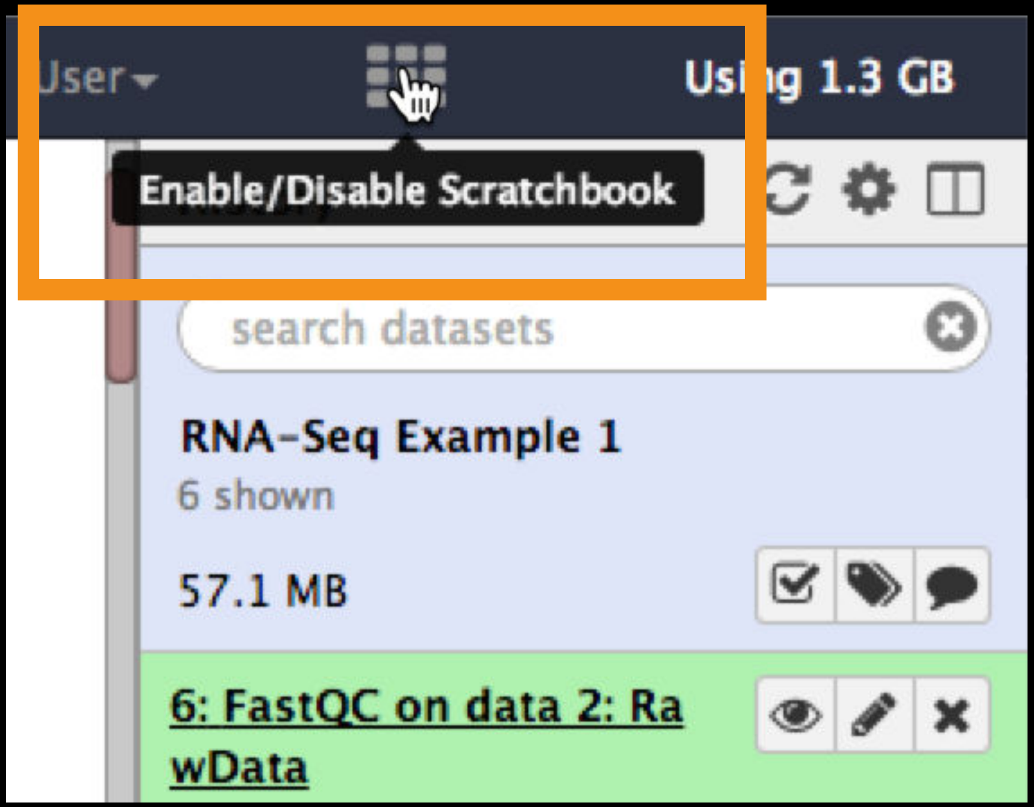
(Count Repeats per exon)

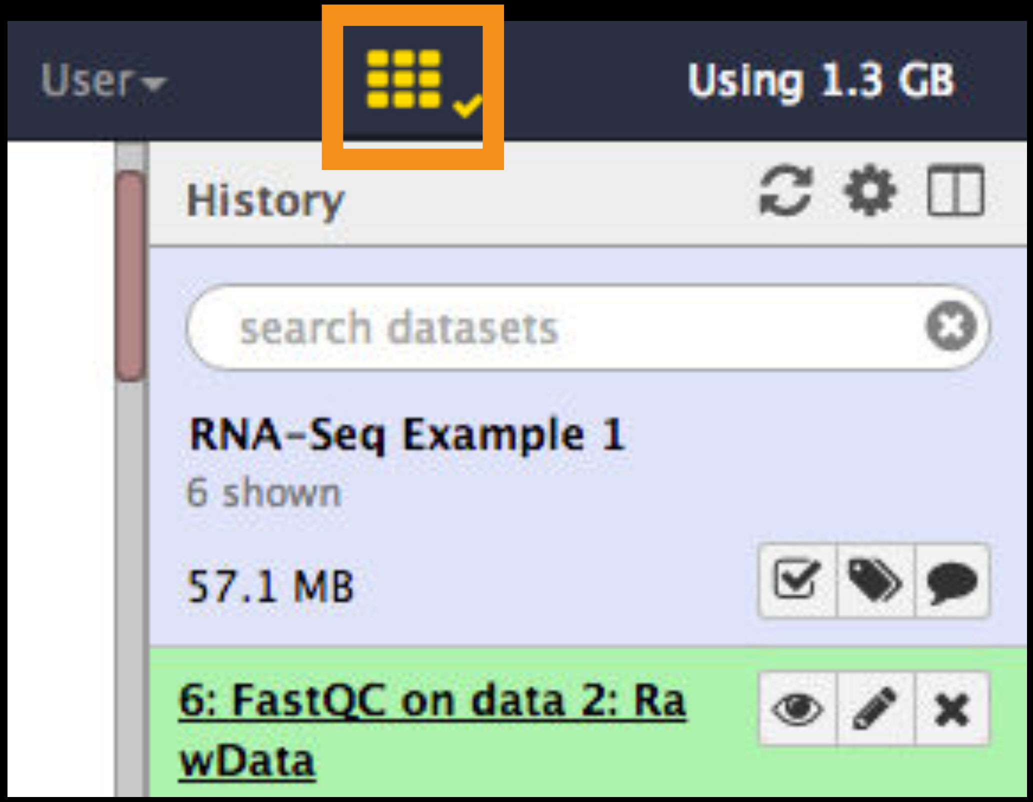
Exon overlap counts

Join, Subtract, and Group → Group

Published History: Exons with overlapping repeats, basic

Scratchbook: View multiple datasets





And the icon turns **yellow**!

Poke a dataset in the eye, and then poke another dataset in the eye.

Yay!

We have exon names and counts!

We are now going to extend that work.

Let's create a copy of this history that we will extend.

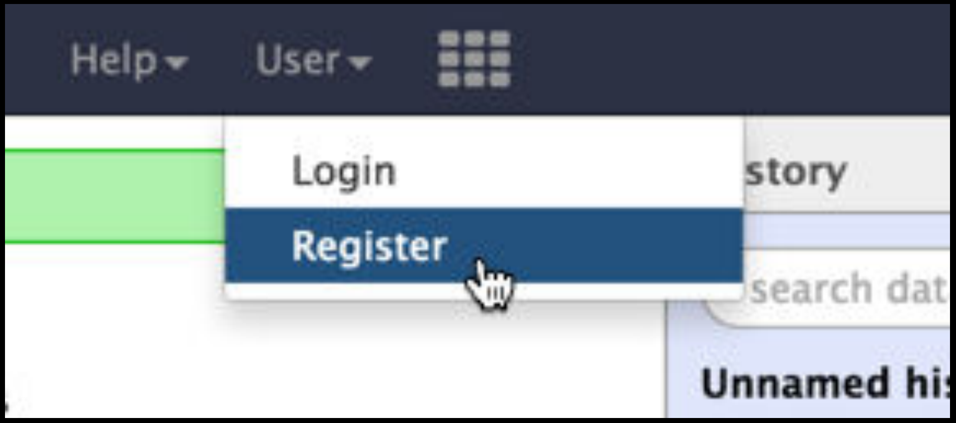
But first, create a login

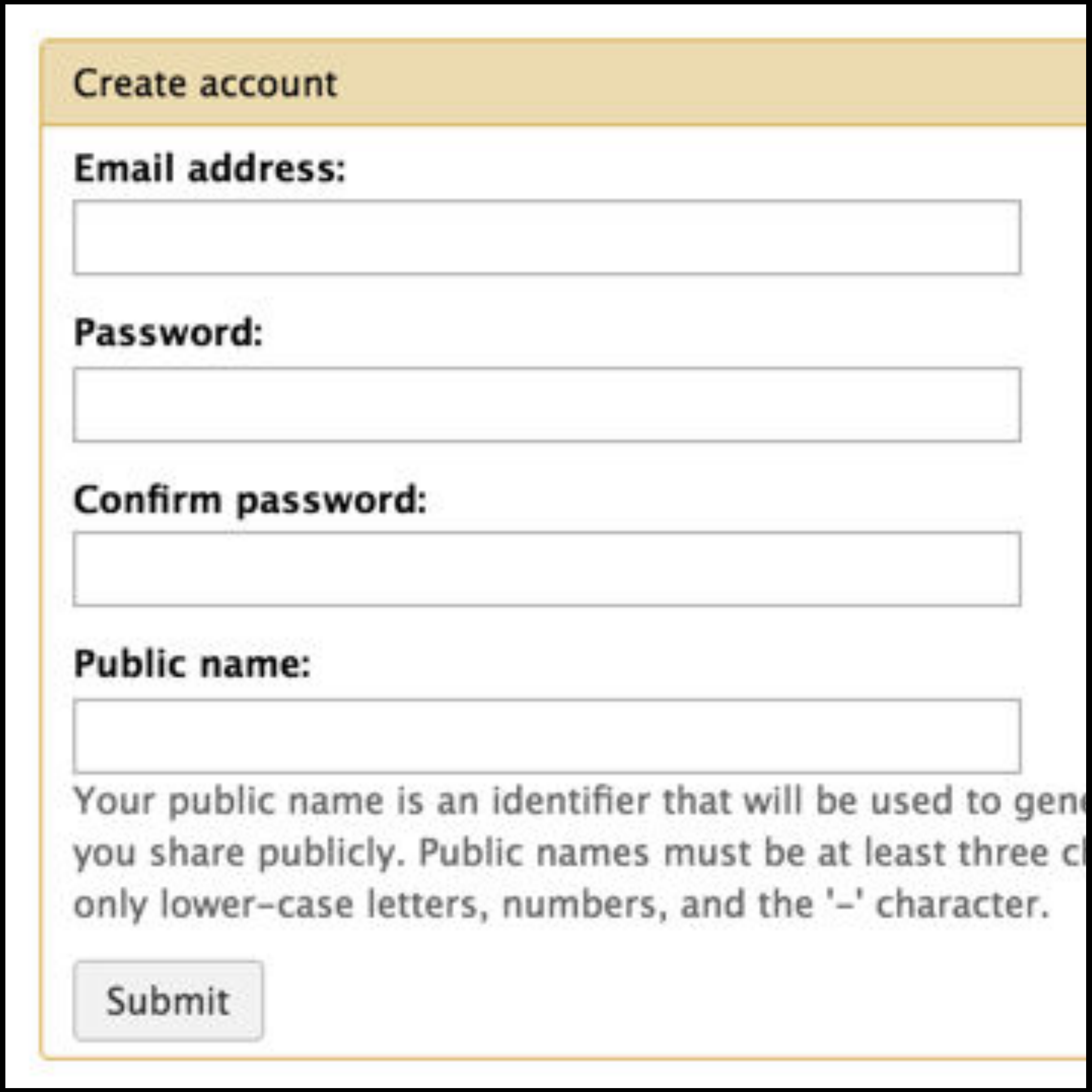
Don't need to login to use Galaxy, but do need one to use all its features

Use an email address you can remember.

Use a low security password.

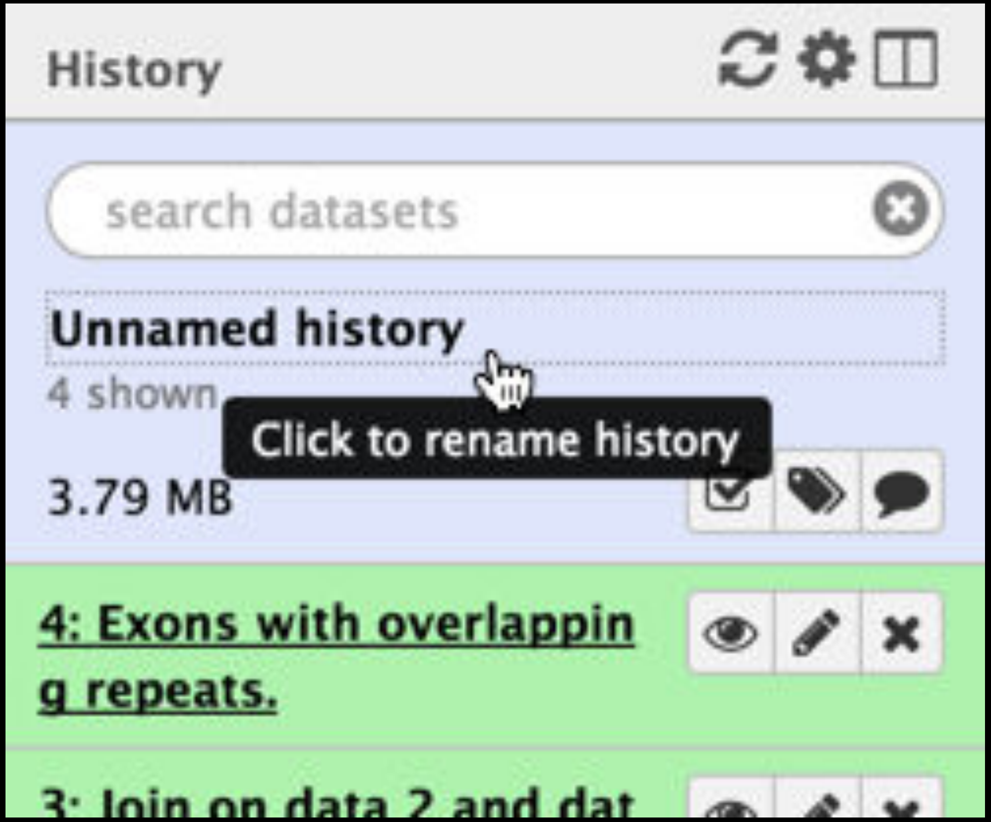
This account will go away after class.





Your public name is an identifier that will be used to gene
you share publicly. Public names must be at least three c
only lower-case letters, numbers, and the '-' character.

Second, name your existing history



History

search datasets

Unnamed history

4 shown

Click to rename history

3.79 MB

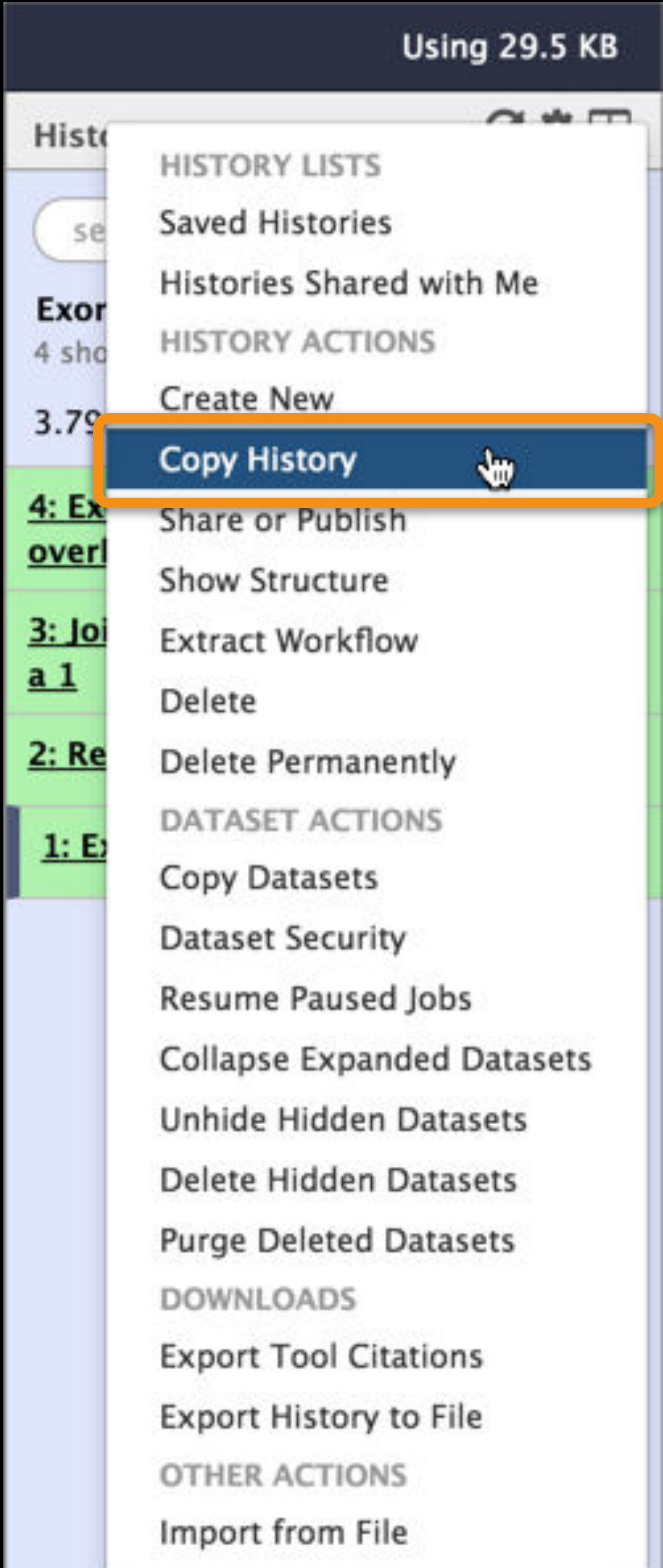
4: Exons with overlapping repeats.

Give your existing history a meaningful name.

3rd, make a copy of your history

(cog) → Copy History
Name the copy based on the exercise you pick

Becomes your new current history.



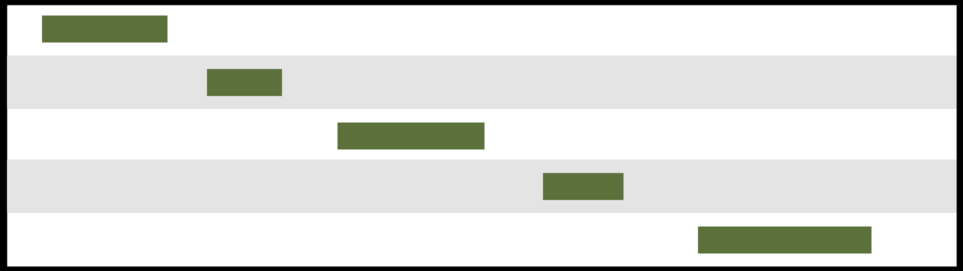
Exons & Repeats: Convert our list to BED format

Create the list of exons with overlapping repeats, in 6-column BED format. Set the score column to be the number of overlapping repeats that exon has.

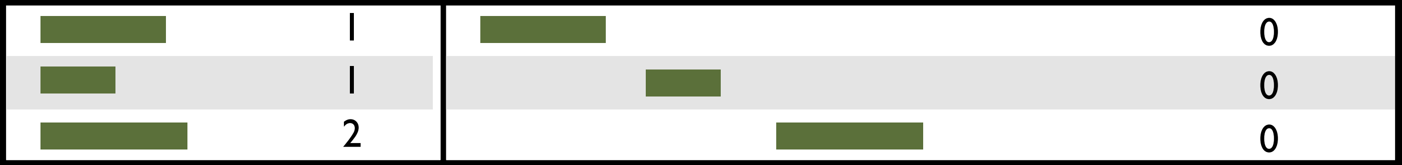
Everything you need will be in these toolboxes

- Text manipulation (cut is particularly useful)
- Join, subtract and group
- Filter and sort
- Operate on genomic intervals

Exon overlap counts



Exons

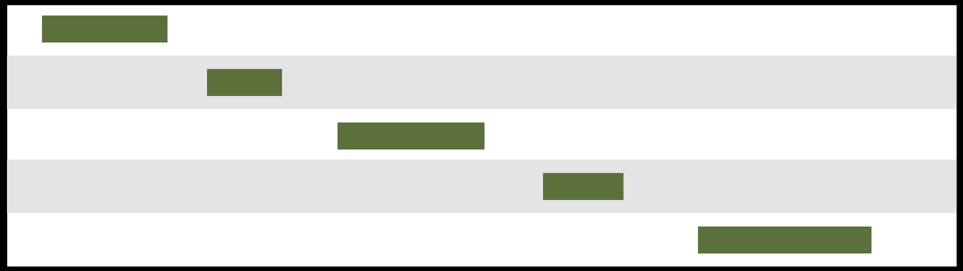


Join on exon name

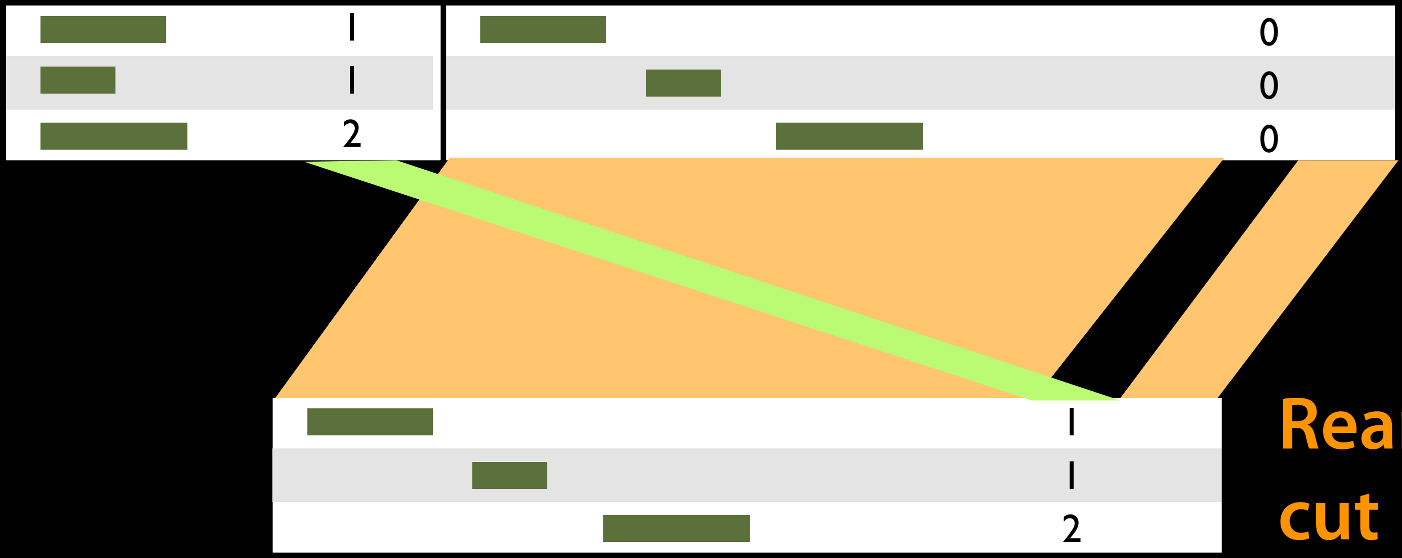
Join, Subtract, and Group → Join

(Incorporate the overlap count with rest of Exon information)

Exon overlap counts



Exons



Join on exon name

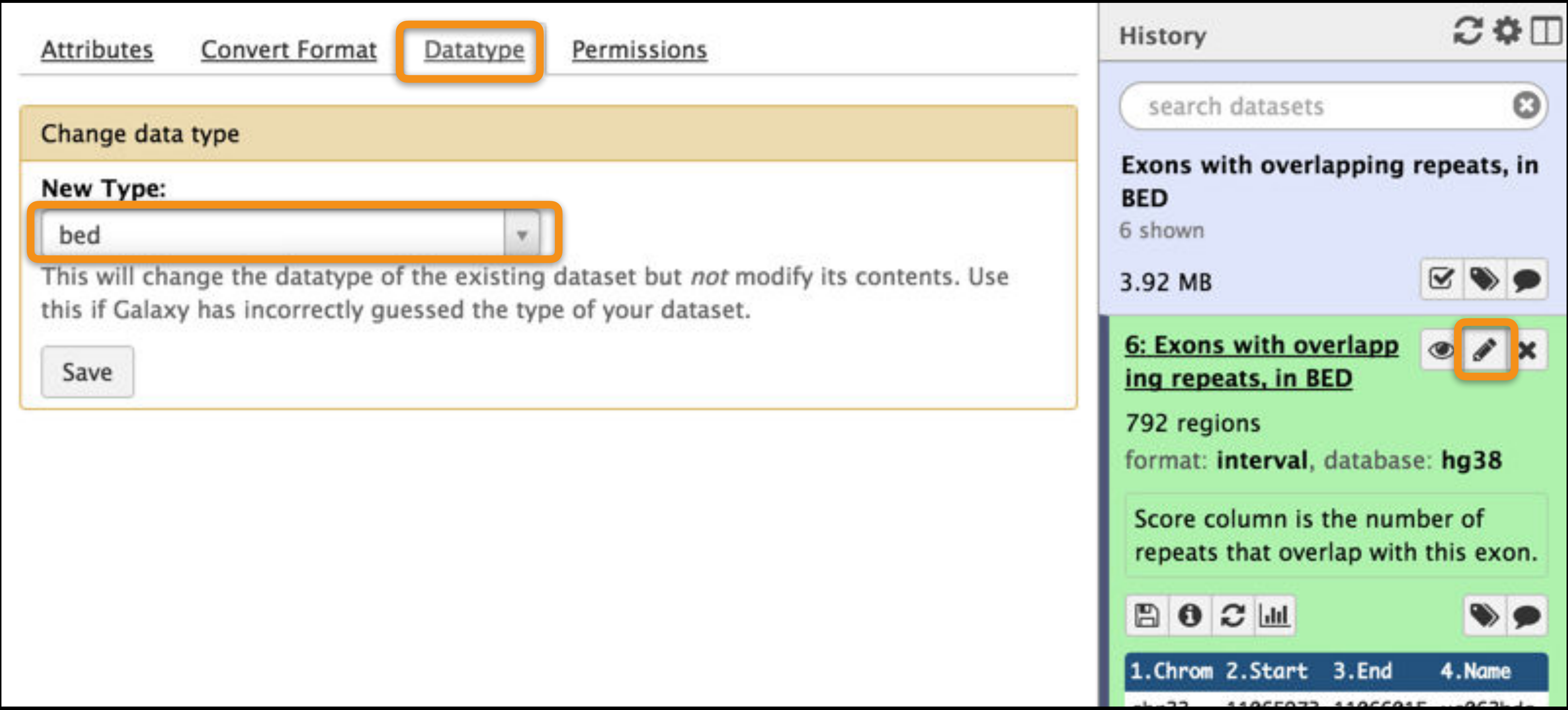
Rearrange columns w/ cut

Text Manipulation → Cut

(Incorporate the overlap count with rest of Exon information)

List of exons with overlaps, in BED format

Edit attributes step at the end:



Change data type

New Type: bed

This will change the datatype of the existing dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Exons with overlapping repeats, in BED

6 shown

3.92 MB

6: Exons with overlapping repeats, in BED

792 regions

format: **interval**, database: **hg38**

Score column is the number of repeats that overlap with this exon.

Published History: Exons with overlapping repeats, in BED

Some Galaxy Terminology

Dataset:

Any input, output or intermediate set of data + metadata

History:

A series of inputs, analysis steps, intermediate datasets, and outputs

Workflow:

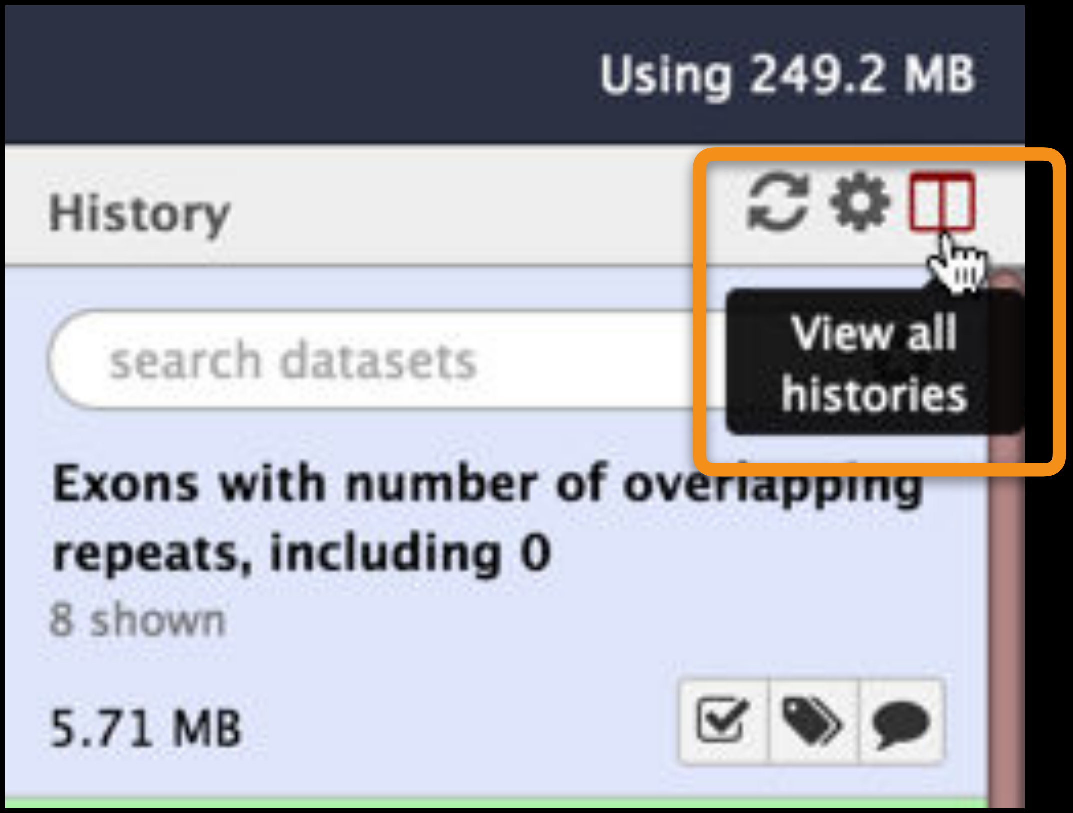
A series of analysis steps

Can be repeated with different data

Exons and Repeats *History* → Reusable *Workflow?*

- The analysis we just finished was about
  - Human chr22
  - Overlap between exons and repeats
  - And then rolling that up to genes
- But, ...
  - is there anything inherent in the analysis about humans, exons or repeats?

Get back to the original history



Using 249.2 MB

History

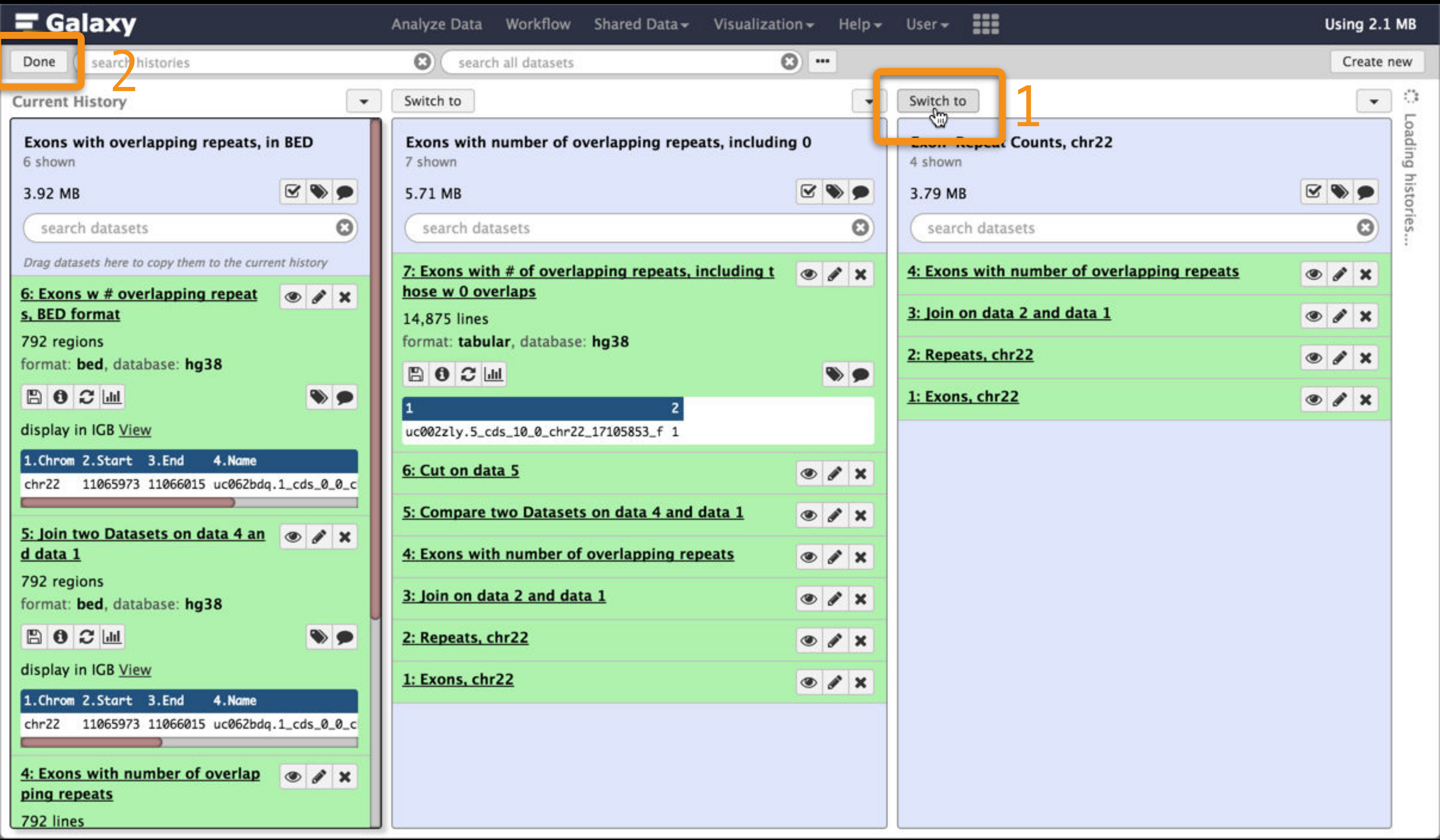
search datasets

Exons with number of overlapping repeats, including 0

8 shown

5.71 MB

Get back to the original history



Exons with overlapping repeats, in BED
6 shown
3.92 MB

Drag datasets here to copy them to the current history

6: Exons w # overlapping repeats, BED format
792 regions
format: bed, database: hg38
display in IGB View

1.Chrom	2.Start	3.End	4.Name
chr22	11065973	11066015	uc062bdq.1_cds_0_0_c

5: Join two Datasets on data 4 and data 1
792 regions
format: bed, database: hg38
display in IGB View

1.Chrom	2.Start	3.End	4.Name
chr22	11065973	11066015	uc062bdq.1_cds_0_0_c

4: Exons with number of overlapping repeats
792 lines

Exons with number of overlapping repeats, including 0
7 shown
5.71 MB

7: Exons with # of overlapping repeats, including those w 0 overlaps
14,875 lines
format: tabular, database: hg38

1	2
uc002zly.5_cds_10_0_chr22_17105853_f	1

6: Cut on data 5
5: Compare two Datasets on data 4 and data 1
4: Exons with number of overlapping repeats
3: Join on data 2 and data 1
2: Repeats, chr22
1: Exons, chr22

Exon Repeat Counts, chr22
4 shown
3.79 MB

4: Exons with number of overlapping repeats
3: Join on data 2 and data 1
2: Repeats, chr22
1: Exons, chr22

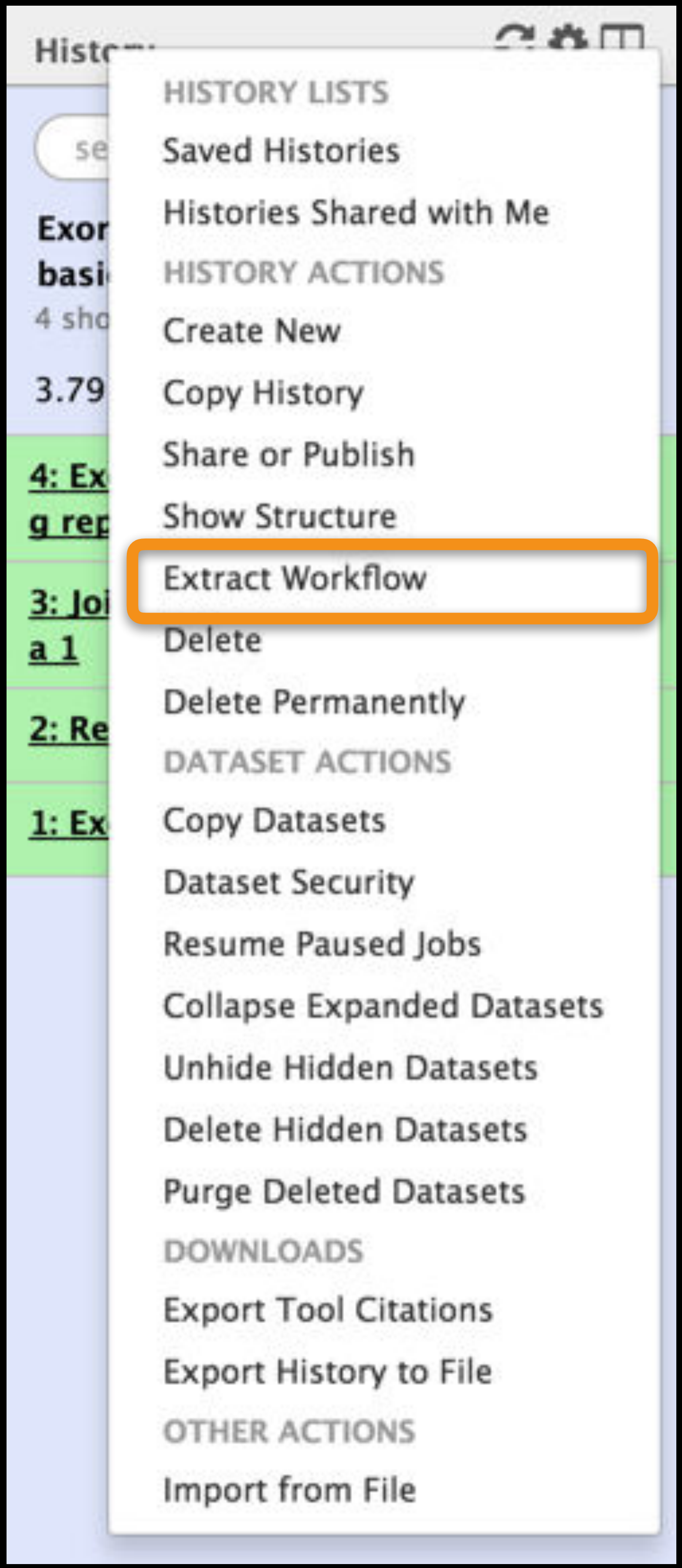
Create a Workflow from a History

Extract Workflow from history

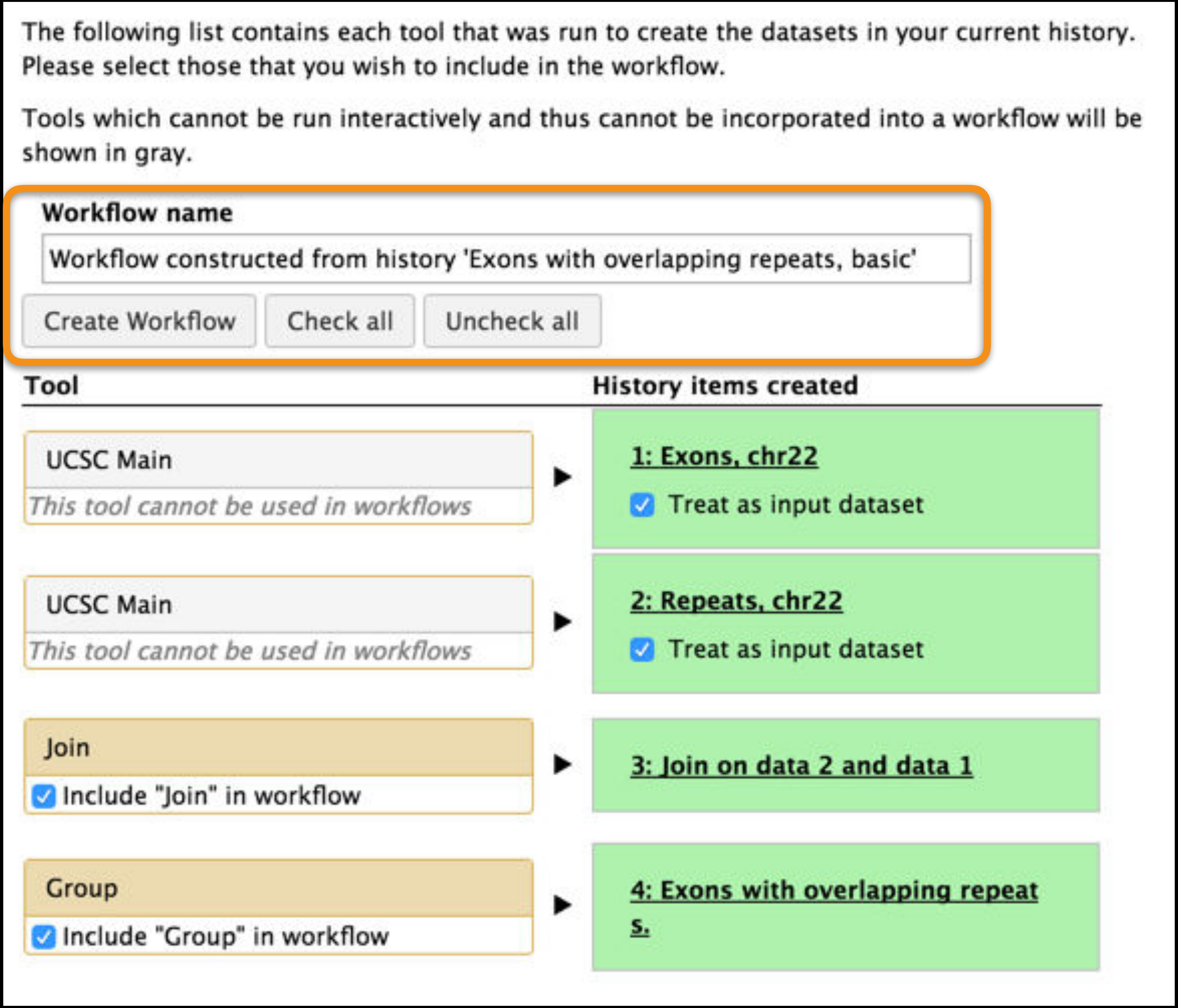
Create a workflow from this history.

Edit it to make some things clearer.

⚙ (cog) → Extract Workflow



Create a Workflow from a History: ...



The following list contains each tool that was run to create the datasets in your current history. Please select those that you wish to include in the workflow.

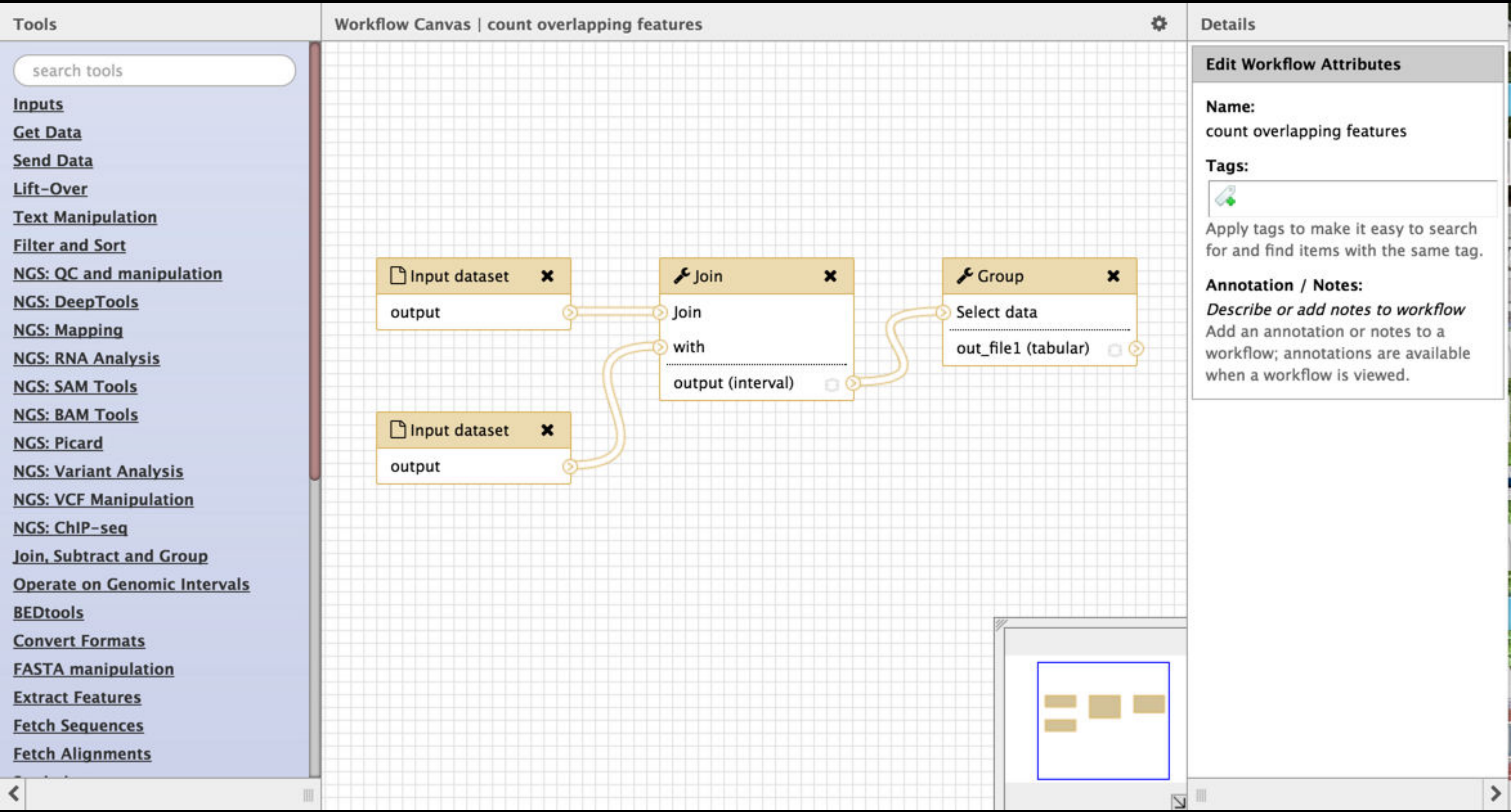
Tools which cannot be run interactively and thus cannot be incorporated into a workflow will be shown in gray.

Workflow name

Workflow constructed from history 'Exons with overlapping repeats, basic'

Tool		History items created
UCSC Main *This tool cannot be used in workflows*	▶	1: Exons, chr22 ☑ Treat as input dataset
UCSC Main *This tool cannot be used in workflows*	▶	2: Repeats, chr22 ☑ Treat as input dataset
Join ☑ Include "Join" in workflow	▶	3: Join on data 2 and data 1
Group ☑ Include "Group" in workflow	▶	4: Exons with overlapping repeats.

Workflow editor



Edit Workflow Attributes

Name:
count overlapping features

Tags:

Apply tags to make it easy to search for and find items with the same tag.

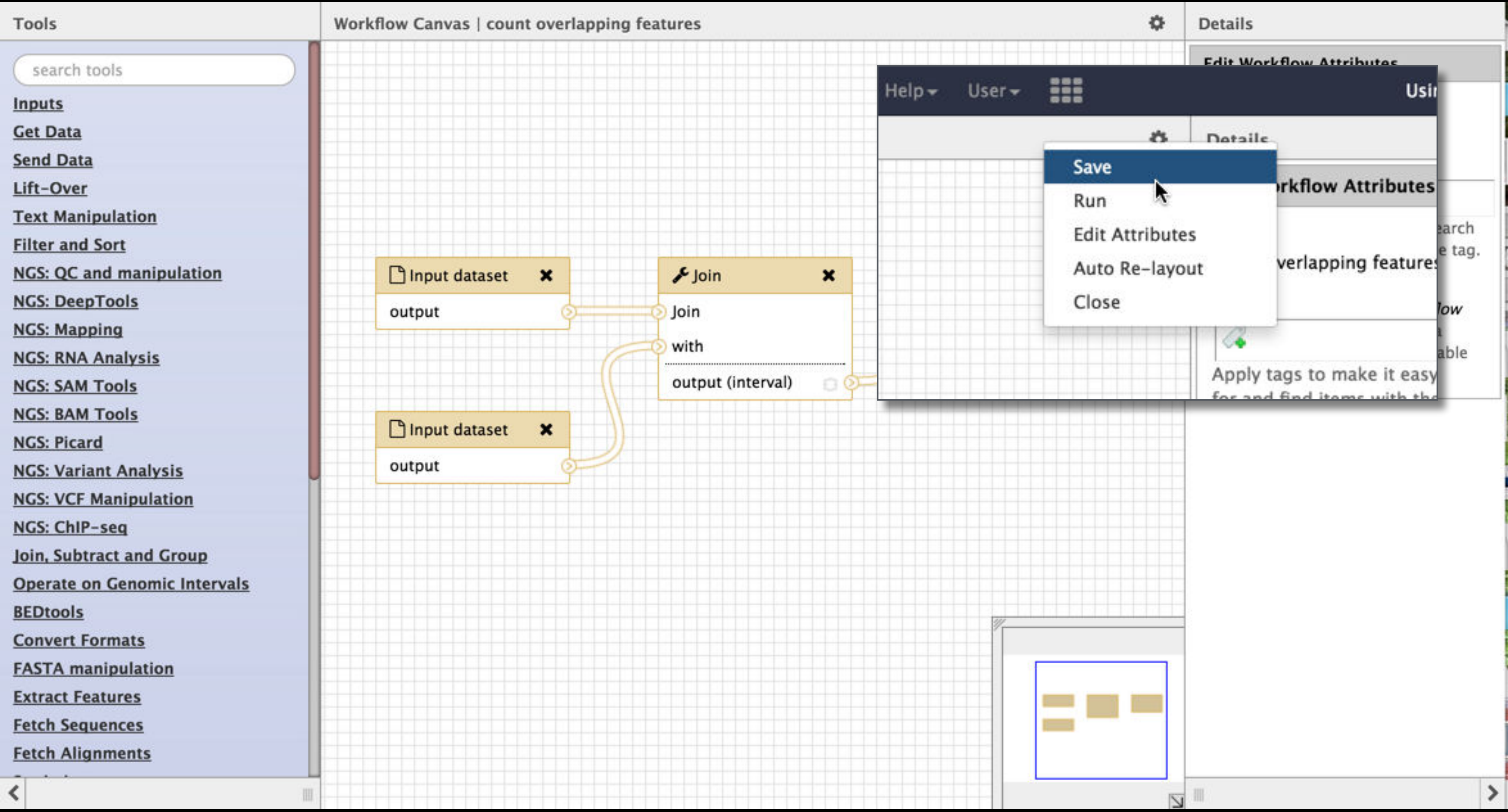
Annotation / Notes:

Describe or add notes to workflow

Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.

Published Workflow: Count Overlaps Between Feature Sets

Workflow editor: save your changes



Published Workflow: Feature Overlap Counting

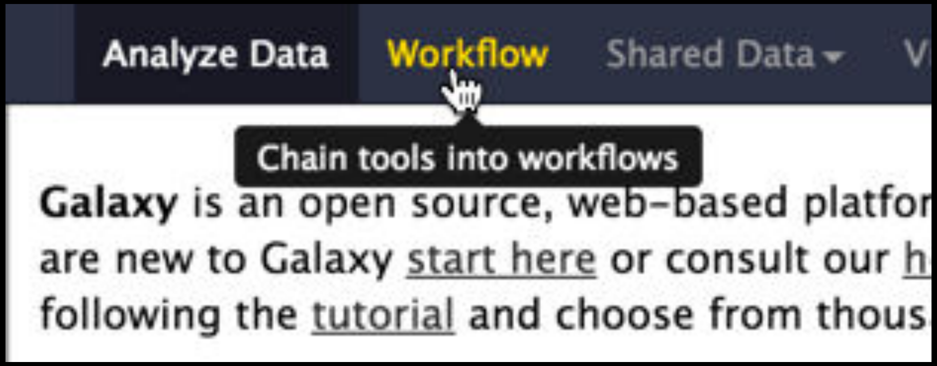
Workflow Testing

Guided: rerun with same inputs

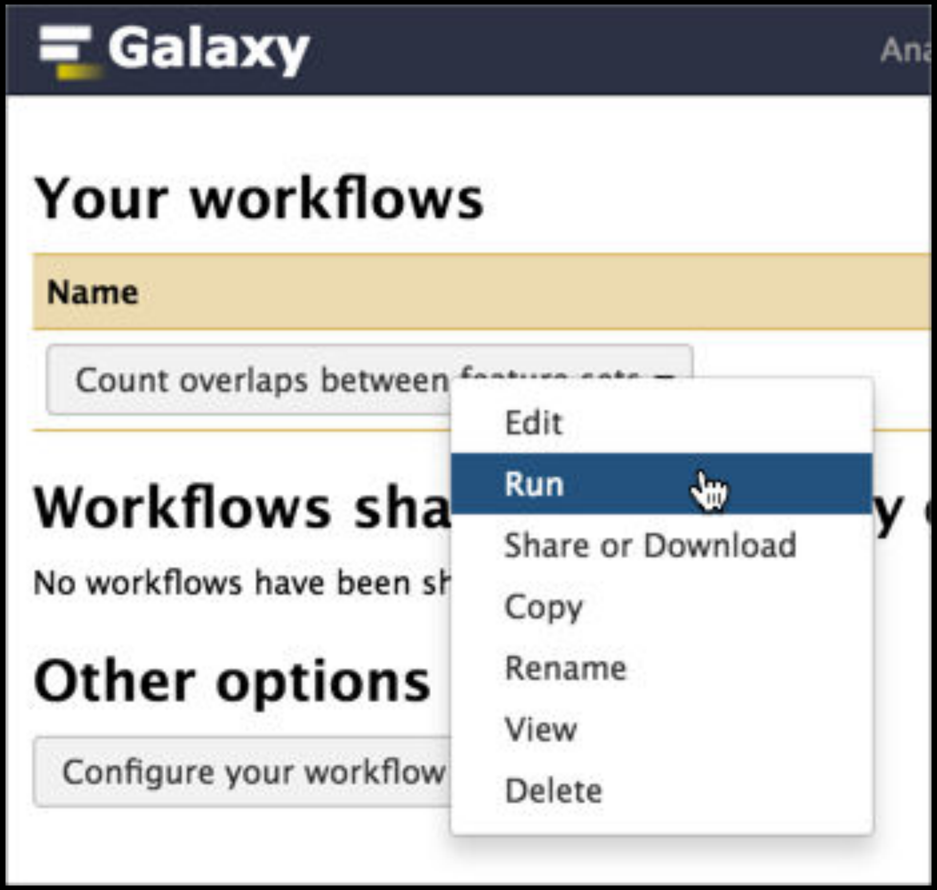
- Workflow → Run
- Did that work?

On your own:

- Count # of exons overlapping each repeat
- Did that work? *Why not?*
- Edit workflow: doc assumptions



Galaxy is an open source, web-based platfor
are new to Galaxy start here or consult our h
following the tutorial and choose from thous



No workflows have been sh

Published Workflow: Count overlaps between feature sets

Workflows: Sweet spots

Short, well-defined tasks, with well-defined inputs and outputs.

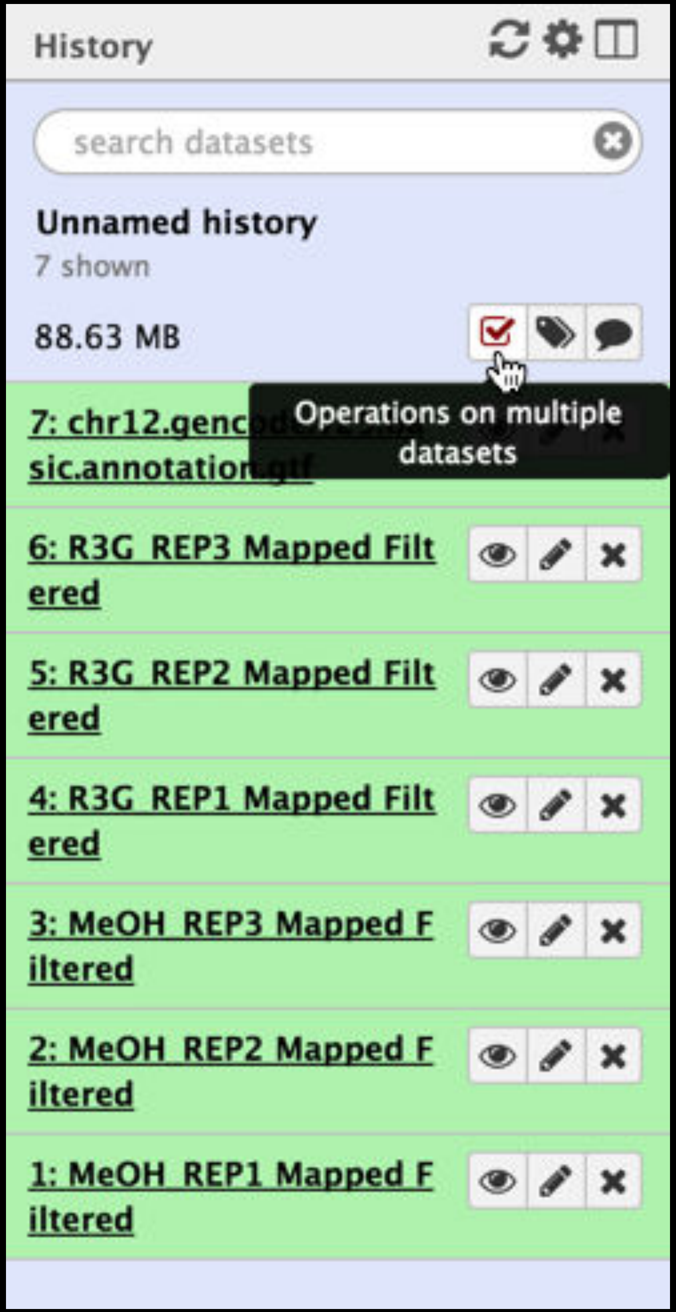
Analysis pipelines for large experiments with many samples where sample and data preparation protocols are the same throughout.

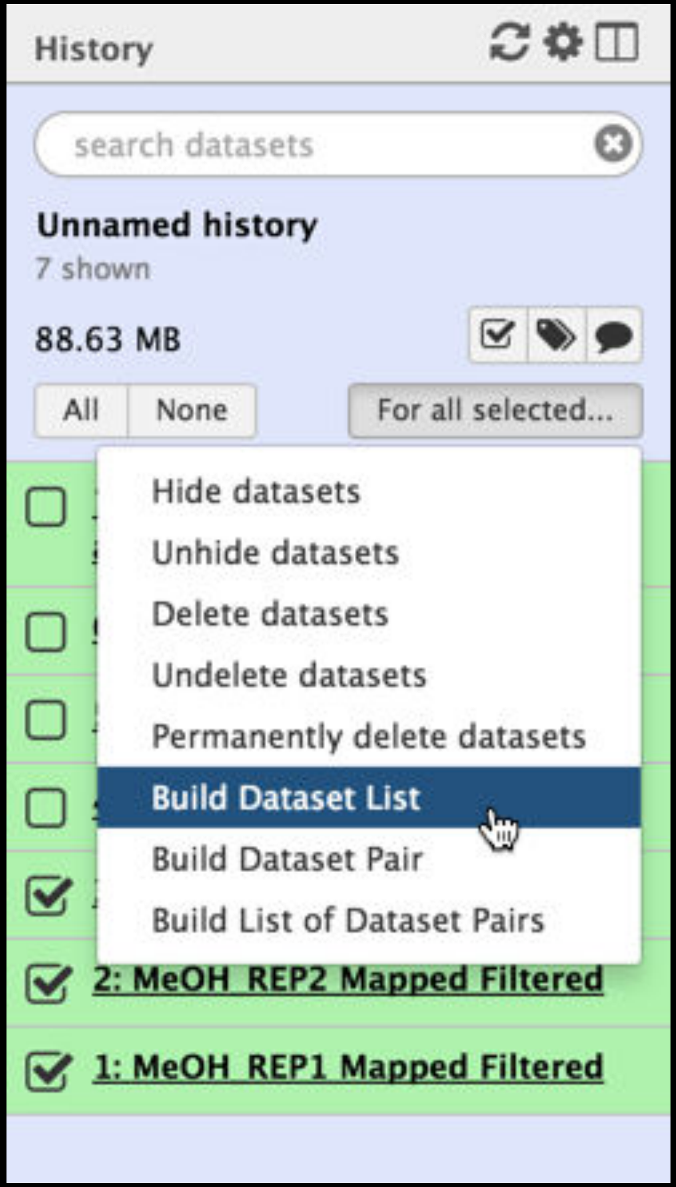
Dataset collections!

Dataset Collections give Galaxy semantic knowledge about dataset relationships.

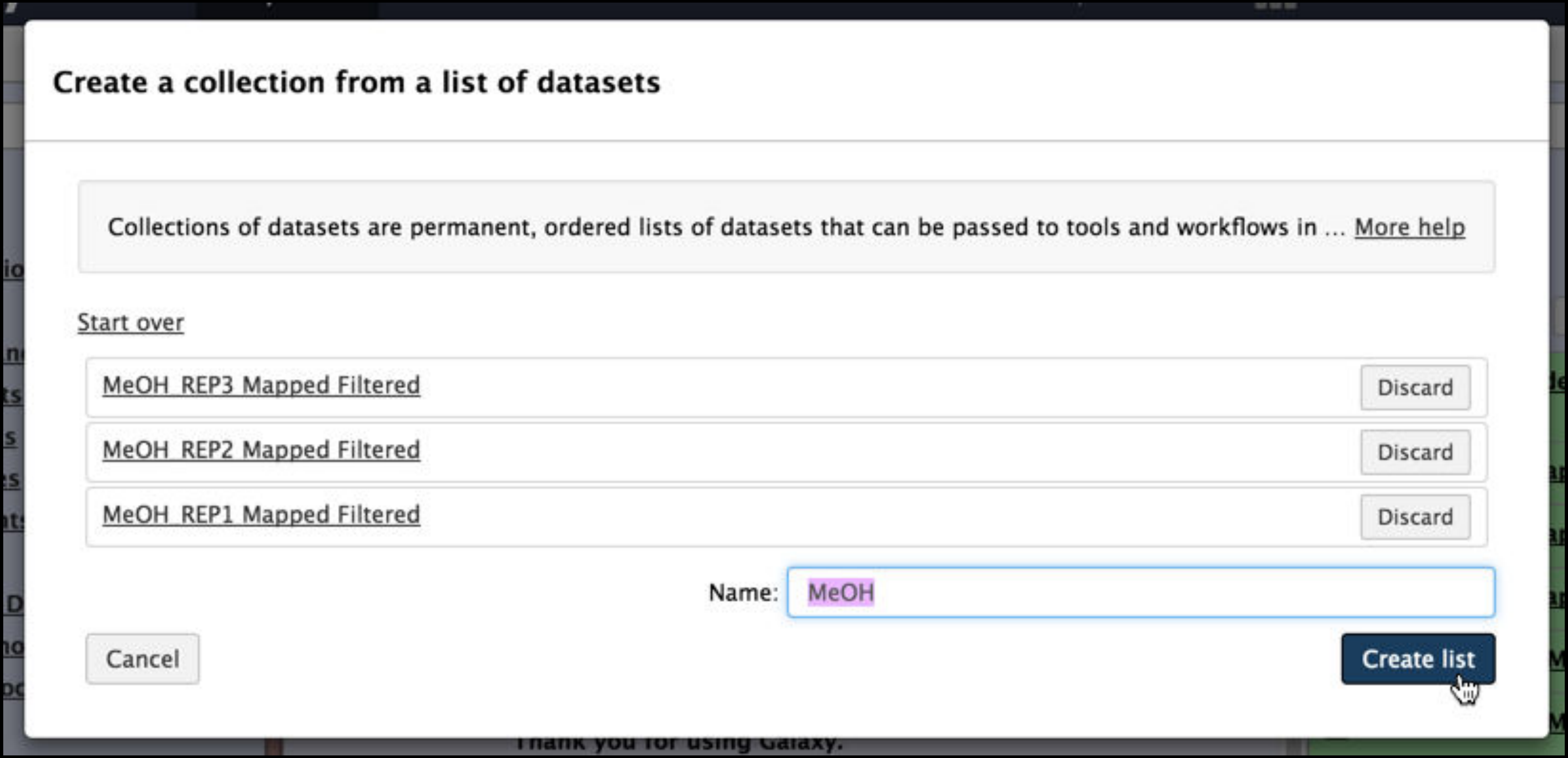
Tools can then take advantage of this knowledge.

Dataset collections





Dataset collections



Create a collection from a list of datasets

Collections of datasets are permanent, ordered lists of datasets that can be passed to tools and workflows in … More help

Start over

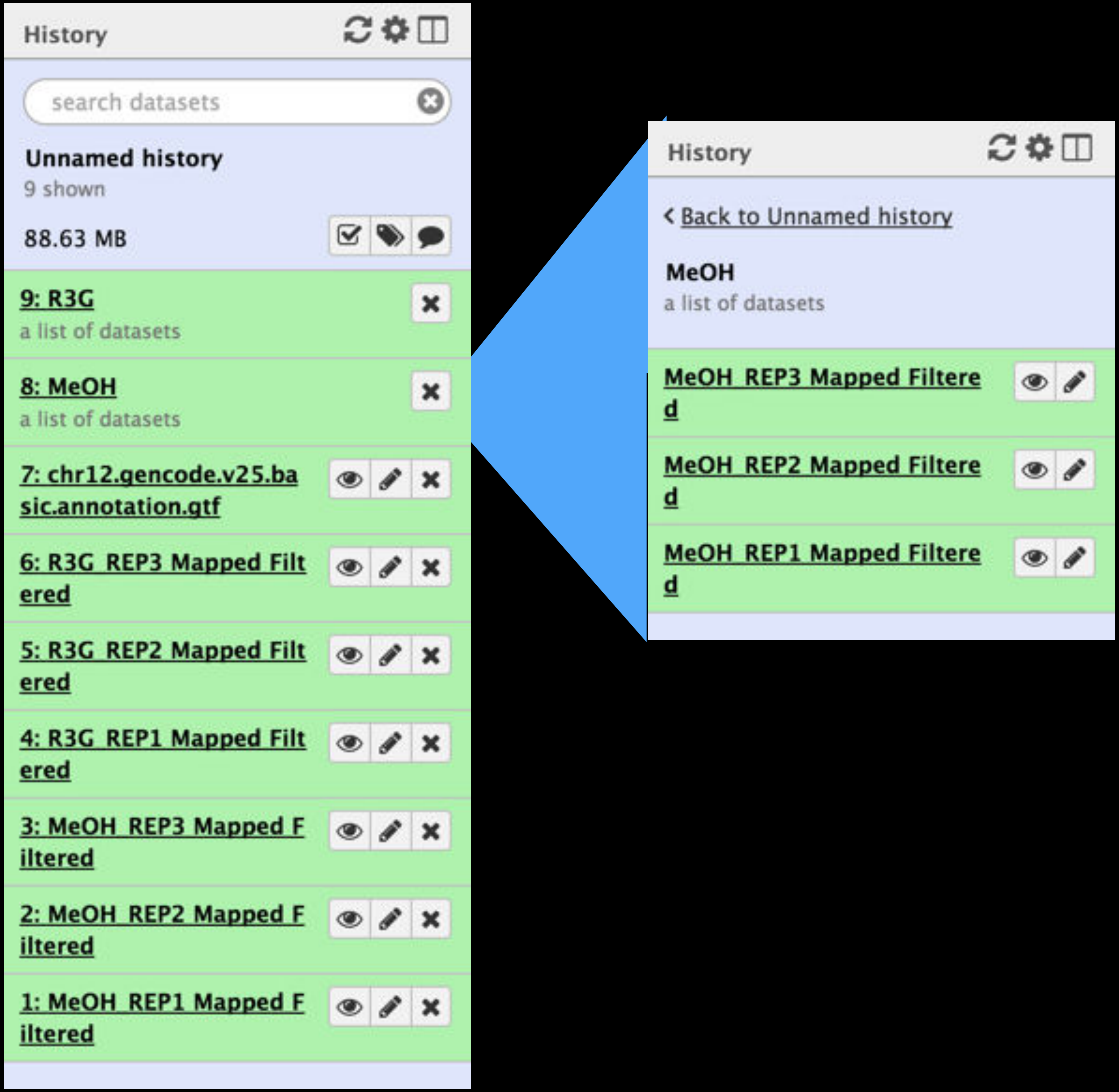
MeOH_REP3 Mapped Filtered

MeOH_REP2 Mapped Filtered

MeOH_REP1 Mapped Filtered

Name: MeOH

Dataset collections



Unnamed history
9 shown

88.63 MB

9: R3G
a list of datasets

8: MeOH
a list of datasets

7: chr12.gencode.v25.basic.annotation.gtf

6: R3G_REP3 Mapped Filtered

5: R3G_REP2 Mapped Filtered

4: R3G_REP1 Mapped Filtered

3: MeOH_REP3 Mapped Filtered

2: MeOH_REP2 Mapped Filtered

1: MeOH_REP1 Mapped Filtered

MeOH
a list of datasets

MeOH_REP3 Mapped Filtered

MeOH_REP2 Mapped Filtered

MeOH_REP1 Mapped Filtered

How to better understand bioinformatics & Galaxy

- **Experiment.** (You are already used to the idea and) Galaxy makes it easy
- **Read** tool documentation and tool and method review papers
- **Get Help!**
  - http://biostars.org/
  - http://seqanswers.com/
  - https://biostar.usegalaxy.org/
  - http://galaxyproject.org/search

SEQanswers
the next generation sequencing community

Biostar
BIOINFORMATICS EXPLAINED

26 - 30 June France

GCC 2017 Montpellier

Le Corum
Conference centre

gcc2017.sciencesconf.org

Galaxy Community Resources: Galaxy Biostar

Tens of thousands of users leads to a lot of questions.

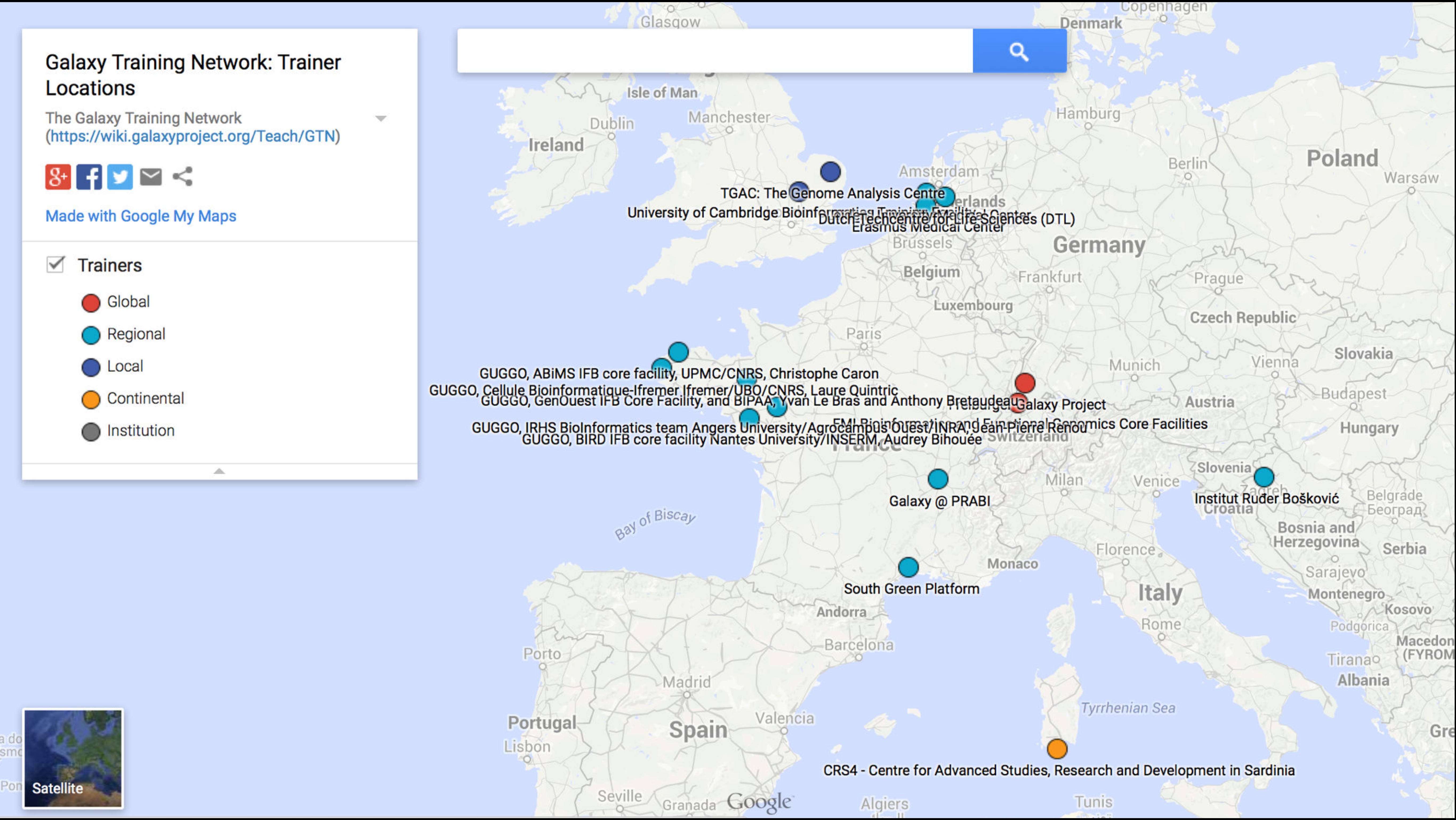
Absolutely have to encourage community support.

Project traditionally used mailing list

Moved the user support list to Galaxy Biostar, an online forum, that uses the Biostar platform

https://biostar.usegalaxy.org/

Scaling Training



Galaxy Training Network: Trainer Locations

The Galaxy Training Network (https://wiki.galaxyproject.org/Teach/GTN)

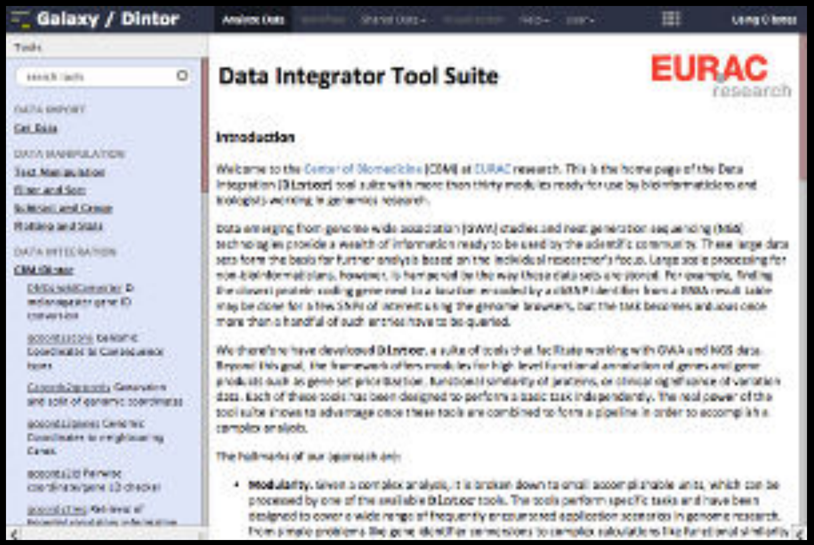
Made with Google My Maps

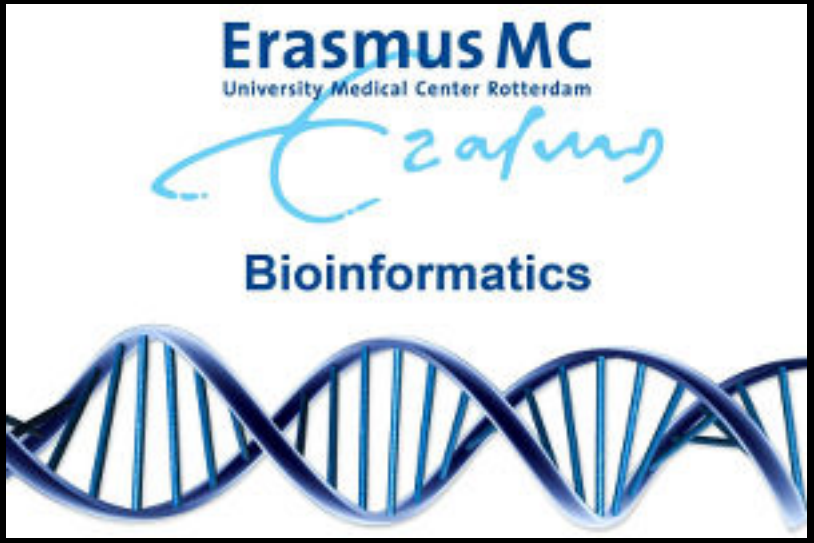
☑ Trainers

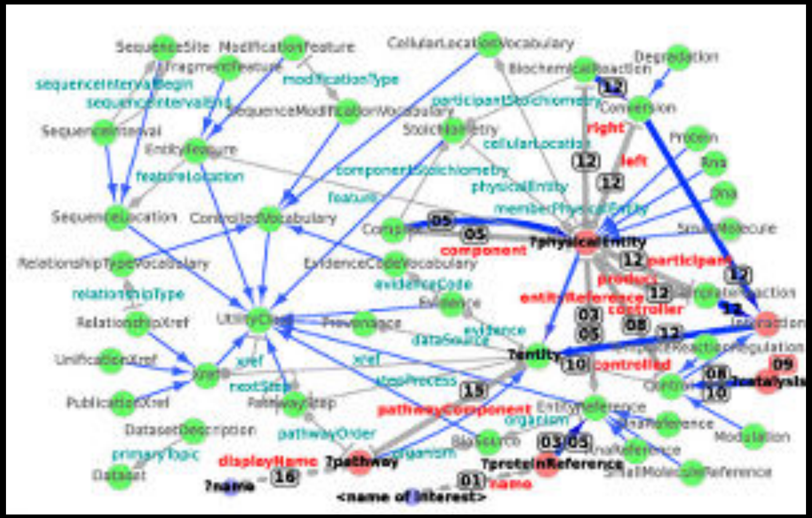
- ● Global
- ● Regional
- ● Local
- ● Continental
- ● Institution

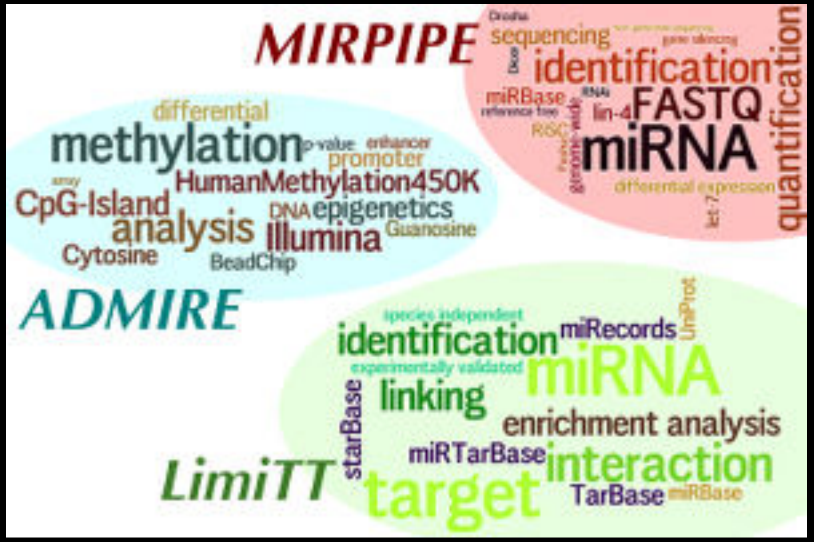
Galaxy Training Network

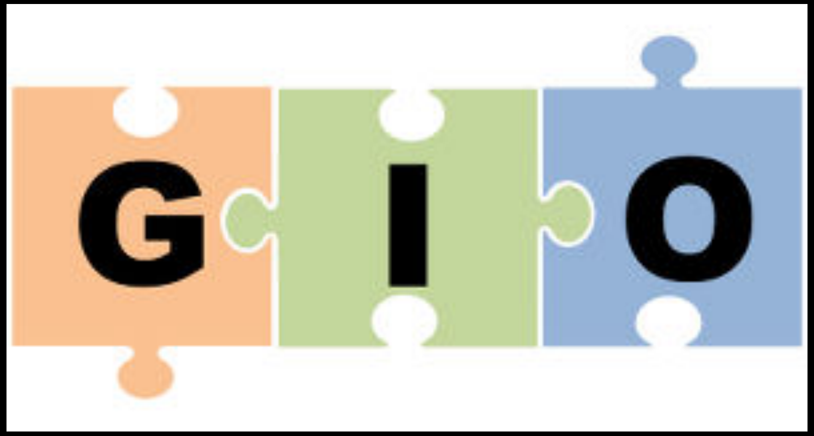
bit.ly/gxygtn

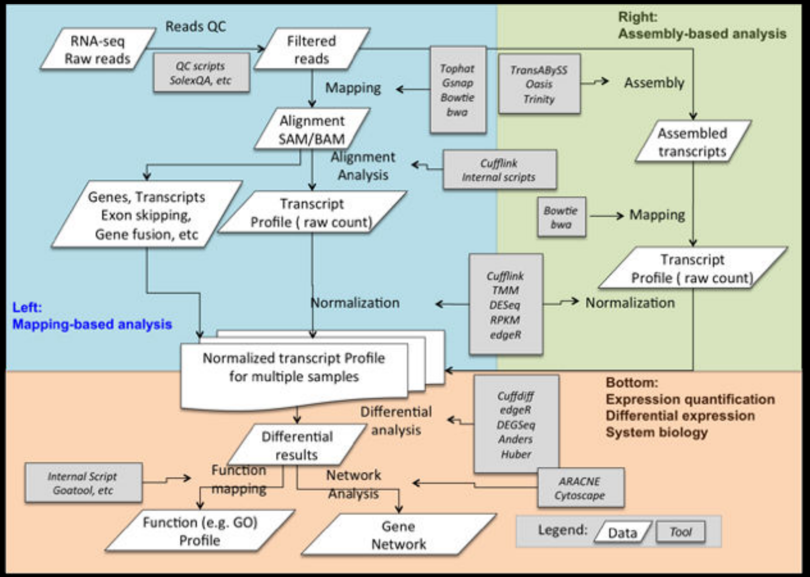


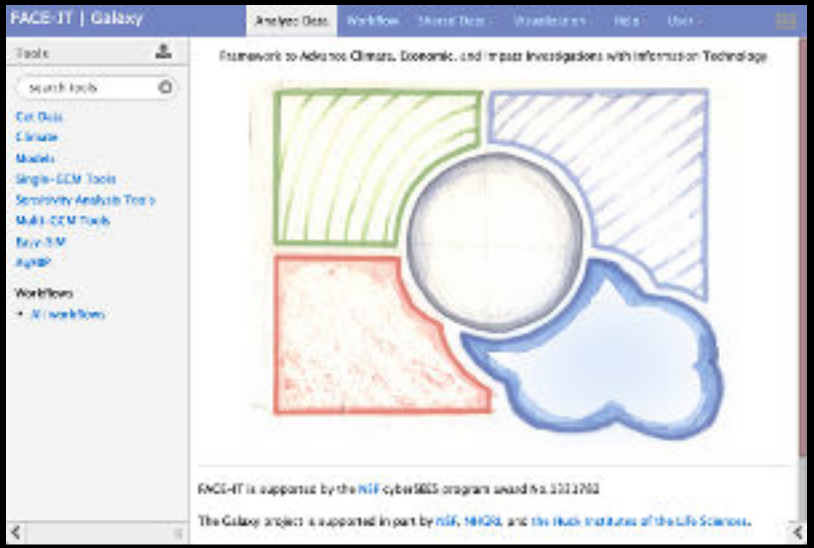








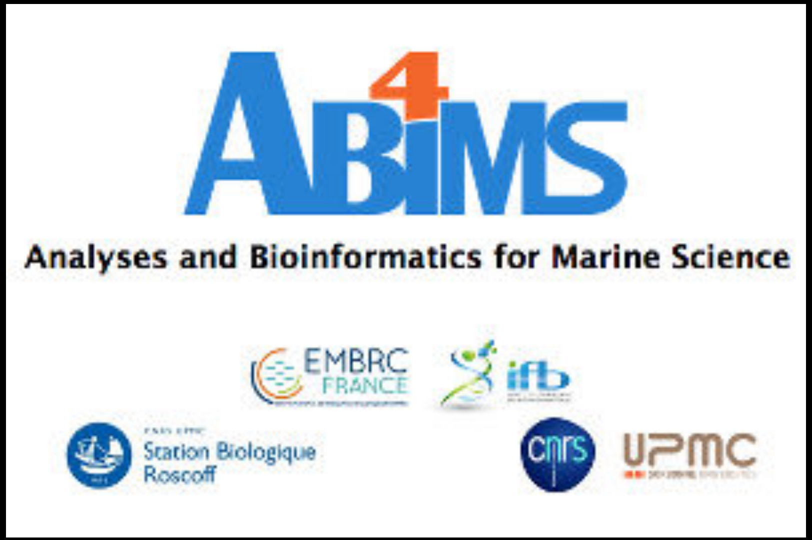




Framework to Advance Climate, Economic, and Impact Investigations with Information Technology

The Galaxy project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

Proteomics
Metabolomics
Natural Language
Image Analysis
Climate Change
Social Science
Cosmology



Galaxy Community Resources: Mailing Lists

http://wiki.galaxyproject.org/MailingLists

Galaxy-Dev

Questions about developing for and deploying Galaxy

High volume (2336 posts in 2015, 1000+ members)

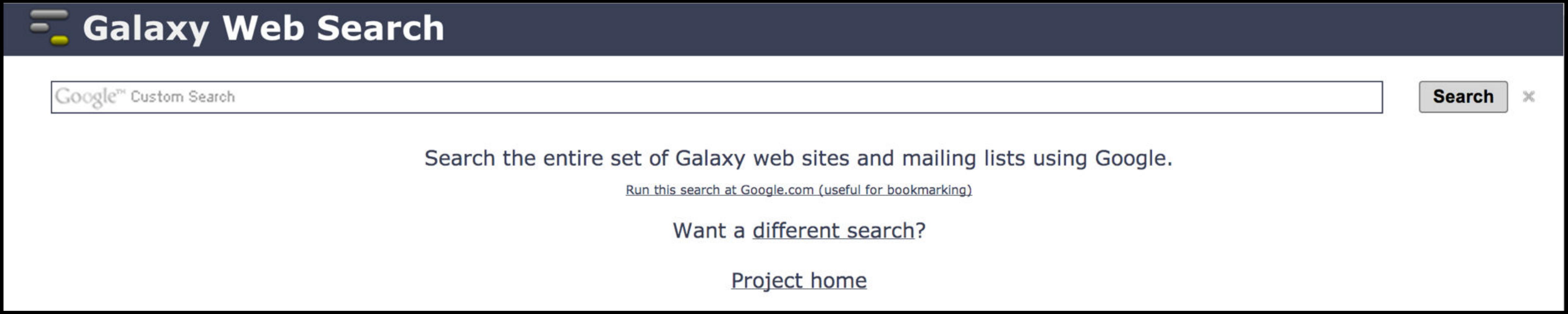
Galaxy-Announce

Project announcements, low volume, moderated

Low volume (36 posts in 2015, 6500+ members)

Also Galaxy-UK, -France, -Proteomics, -Training, ...

Unified Search: http://galaxyproject.org/search



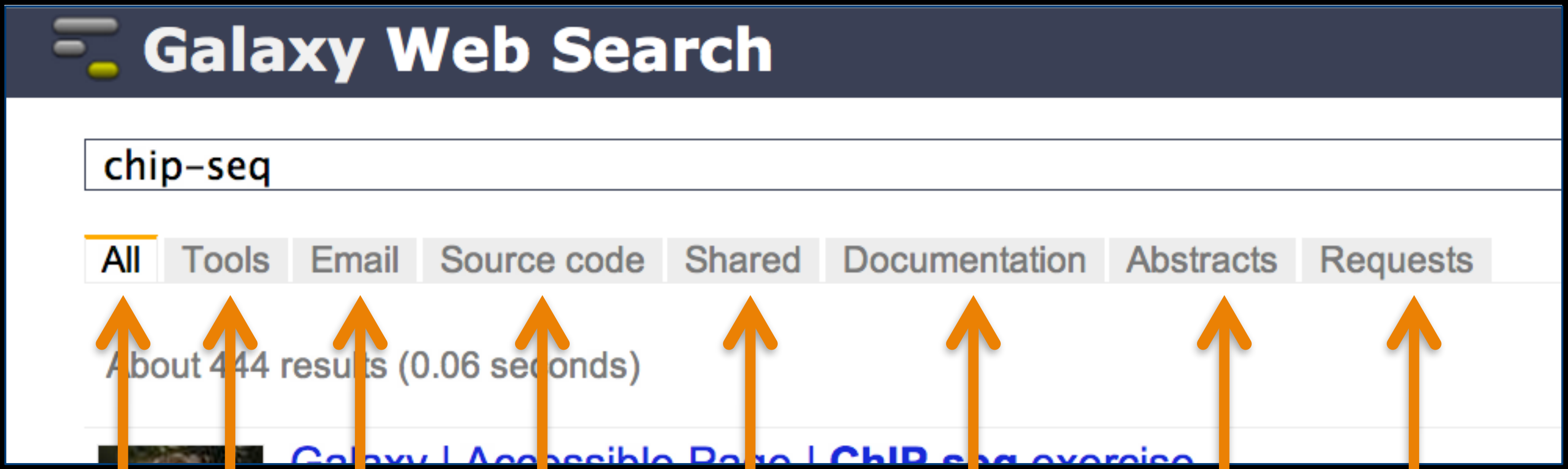
Galaxy Web Search

Search the entire set of Galaxy web sites and mailing lists using Google.

Run this search at Google.com (useful for bookmarking)

Want a different search?

Project home



Galaxy Web Search

chip-seq

About 444 results (0.06 seconds)

Find

Everything on …

Tools for …

Email about …

Source code for …

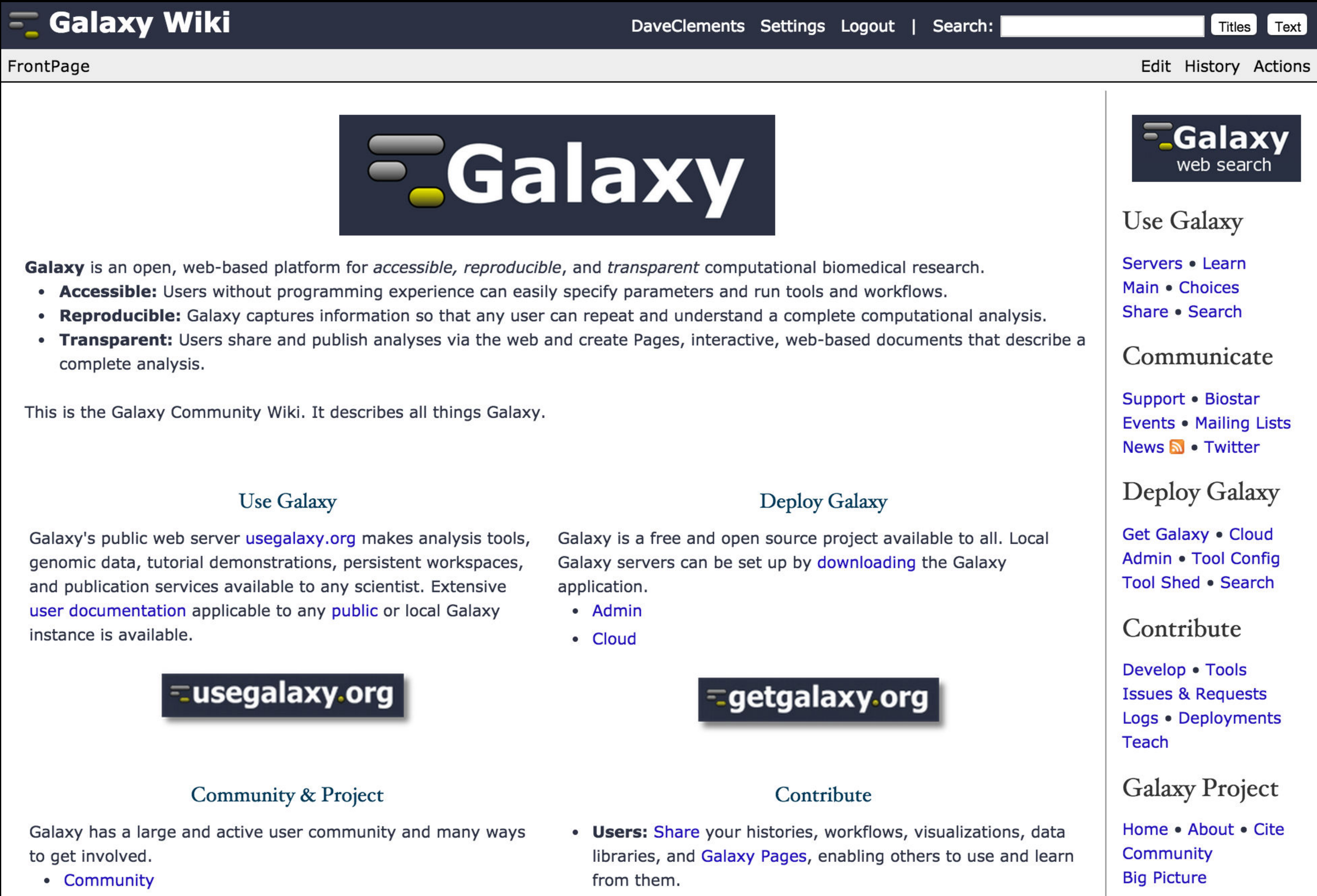
Published Histories, Pages, Workflows, about …

Documentation on …

Papers using Galaxy for …

Related feature requests

http://wiki.galaxyproject.org



Galaxy is an open, web-based platform for *accessible, reproducible*, and *transparent* computational biomedical research.

- **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

This is the Galaxy Community Wiki. It describes all things Galaxy.

Use Galaxy

Galaxy's public web server usegalaxy.org makes analysis tools, genomic data, tutorial demonstrations, persistent workspaces, and publication services available to any scientist. Extensive user documentation applicable to any public or local Galaxy instance is available.

Deploy Galaxy

Galaxy is a free and open source project available to all. Local Galaxy servers can be set up by downloading the Galaxy application.

- Admin
- Cloud

Community & Project

Galaxy has a large and active user community and many ways to get involved.

- Community

Contribute

- **Users:** Share your histories, workflows, visualizations, data libraries, and Galaxy Pages, enabling others to use and learn from them.

Use Galaxy

Servers • Learn
Main • Choices
Share • Search

Communicate

Support • Biostar
Events • Mailing Lists
News • Twitter

Deploy Galaxy

Get Galaxy • Cloud
Admin • Tool Config
Tool Shed • Search

Contribute

Develop • Tools
Issues & Requests
Logs • Deployments
Teach

Galaxy Project

Home • About • Cite
Community
Big Picture

Events

News

Galaxy Event Horizon

Events with Galaxy-related content are listed here.

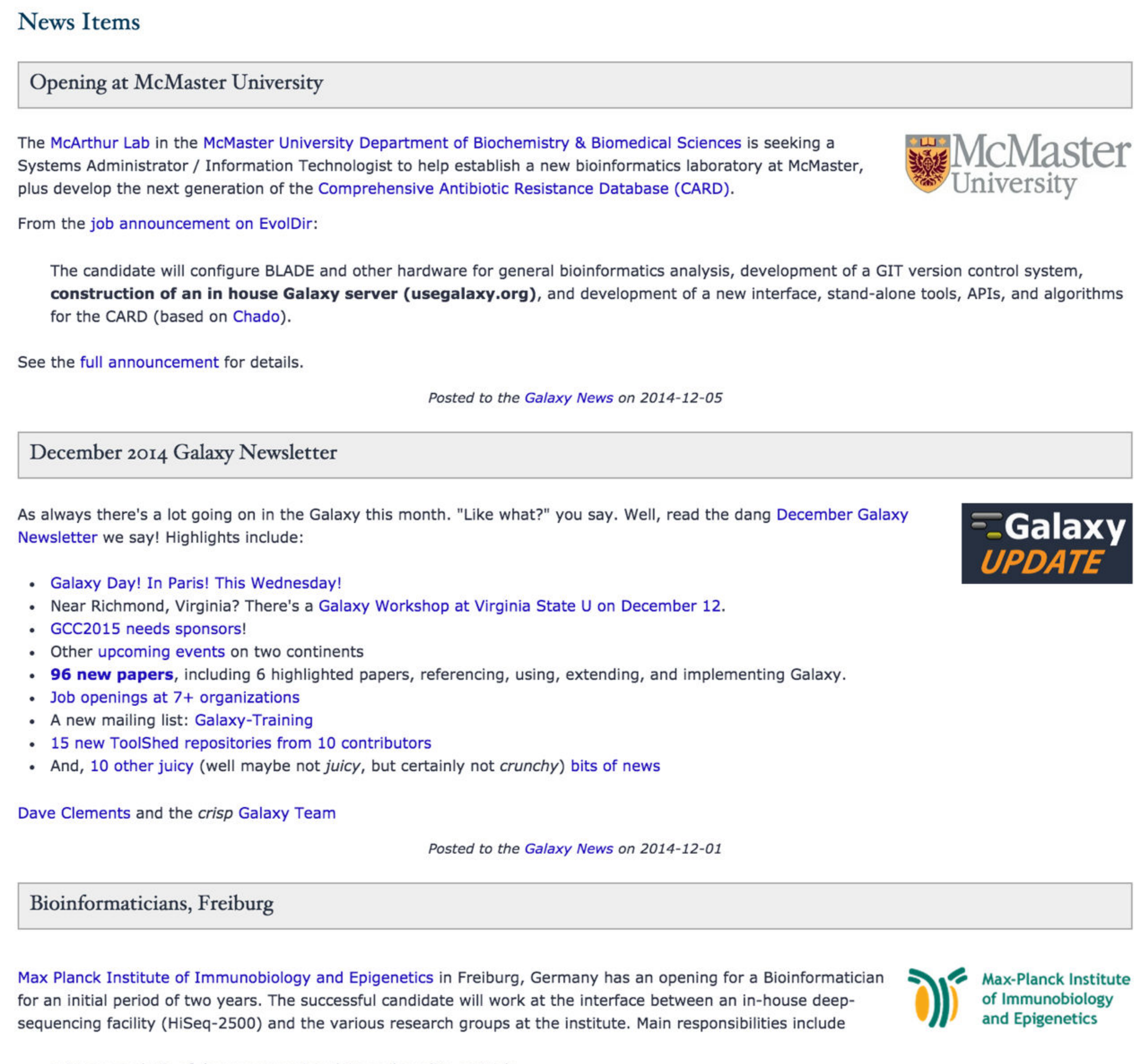
Also see the Galaxy Events Google Calendar for a listing of events and deadlines that are
Galaxy Community. This is also available as an RSS feed.

If you know of any event that should be added to this page and/or to the Galaxy Event Calendar,
send it to outreach@glaxyproject.org.

For events prior to this year, see the **Events Archive**.

Upcoming Events

Date	Topic/Event	Venue/Location
December 12	*Introduction to Galaxy Workshop*	Virginia State University, Petersburg, Virgin
December 16-19	*RNA-Seq and ChIP-Seq Analysis with Galaxy*	UC Davis, California, United States
	2015	
January 10-14	*Galaxy for SNP and Variant Data Analysis*	Plant and Animal Genome XXIII (PAG2014), States
January 19-20	*NGS pipelines with Galaxy*	e-Infrastructures for Massively Parallel Sequ Sweden
February 9-13	*Analyse bioinformatique de séquences sous Galaxy*	Montpellier, France
February 16-18	*Accessible and Reproducible Large-Scale Analysis with Galaxy*	Genome and Transcriptome Analysis, pa Conference, San Francisco, Cali
	Large-Scale NGS data Analysis on Amazon Web Services Using Globus Genomic	Genomics & Sequencing Data Integration, of Molecular Medicine Tri-Conference, Sa States
	iReport: An Integrative "omics"	



News Items

Opening at McMaster University

The McArthur Lab in the McMaster University Department of Biochemistry & Biomedical Sciences is seeking a Systems Administrator / Information Technologist to help establish a new bioinformatics laboratory at McMaster, plus develop the next generation of the Comprehensive Antibiotic Resistance Database (CARD).

From the job announcement on EvolDir:

> The candidate will configure BLADE and other hardware for general bioinformatics analysis, development of a GIT version control system, **construction of an in house Galaxy server (usegalaxy.org)**, and development of a new interface, stand-alone tools, APIs, and algorithms for the CARD (based on Chado).

See the full announcement for details.

Posted to the Galaxy News on 2014-12-05

December 2014 Galaxy Newsletter

As always there's a lot going on in the Galaxy this month. "Like what?" you say. Well, read the dang December Galaxy Newsletter we say! Highlights include:

- Galaxy Day! In Paris! This Wednesday!
- Near Richmond, Virginia? There's a Galaxy Workshop at Virginia State U on December 12.
- GCC2015 needs sponsors!
- Other upcoming events on two continents
- **96 new papers**, including 6 highlighted papers, referencing, using, extending, and implementing Galaxy.
- Job openings at 7+ organizations
- A new mailing list: Galaxy-Training
- 15 new ToolShed repositories from 10 contributors
- And, 10 other juicy (well maybe not *juicy*, but certainly not *crunchy*) bits of news

Dave Clements and the *crisp* Galaxy Team

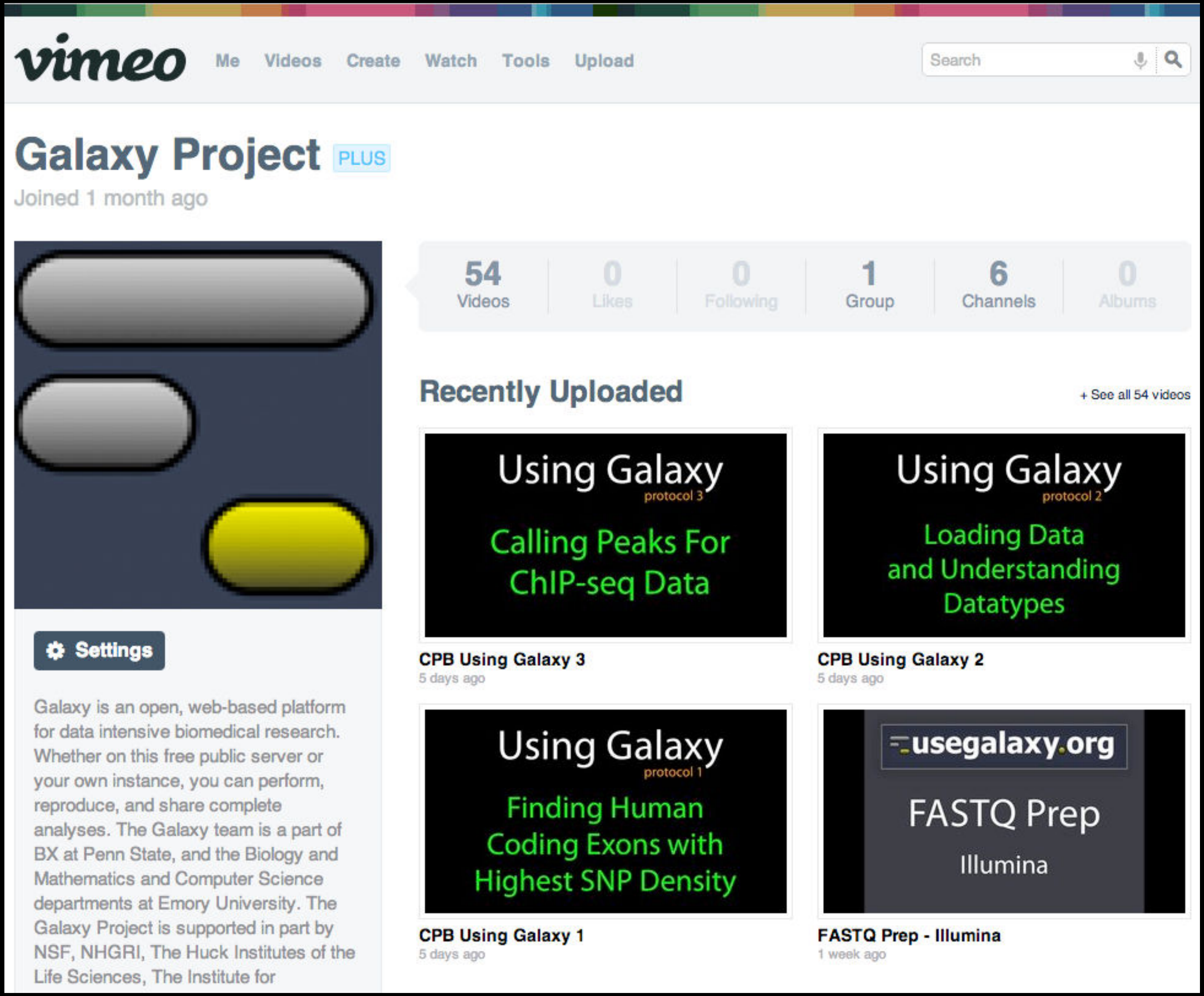
Posted to the Galaxy News on 2014-12-01

Bioinformaticians, Freiburg

Max Planck Institute of Immunobiology and Epigenetics in Freiburg, Germany has an opening for a Bioinformatician for an initial period of two years. The successful candidate will work at the interface between an in-house deep-sequencing facility (HiSeq-2500) and the various research groups at the institute. Main responsibilities include

- primary analysis of deep-sequencing data and quality controls

Galaxy Resources & Community: Videos



Galaxy Project PLUS

Joined 1 month ago

Galaxy is an open, web-based platform for data intensive biomedical research. Whether on this free public server or your own instance, you can perform, reproduce, and share complete analyses. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for

Recently Uploaded

Using Galaxy protocol 3 – Calling Peaks For ChIP-seq Data
CPB Using Galaxy 3
5 days ago

Using Galaxy protocol 2 – Loading Data and Understanding Datatypes
CPB Using Galaxy 2
5 days ago

Using Galaxy protocol 1 – Finding Human Coding Exons with Highest SNP Density
CPB Using Galaxy 1
5 days ago

usegalaxy.org – FASTQ Prep – Illumina
FASTQ Prep - Illumina
1 week ago

"How to" screencasts on using and deploying Galaxy

Talks from previous meetings.

http://vimeo.com/galaxyproject

Galaxy Resources & Community: CiteULike Group

Group: Galaxy - library 3726 articles

You are an administrative member of this group.
Invite other CiteULike users to join, or invite people who don't use CiteULike yet.

✔ **Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a**
eLife, Vol. 5 (25 August 2016), doi:10.7554/elife.19276
by Matthew J. Shurtleff, Morayma M. Temoche-Diaz, Kate V. Karfilis, Sayaka Ri, Randy Sch
posted to methods usemain by galaxyproject to the group Galaxy keyed Shurtleff2016Ybox
■ Copy ■ My Copy

Validation and characterization of thirteen microsatellite markers for queen con
PeerJ Preprints, Vol. 4 (October 2016), doi:10.7287/peerj.preprints.2559v1
by Nathan K. Truelove, Loong Fai Ho, Richard F. Preziosi, Stephen J. Box
posted to methods uselocal by galaxyproject to the group Galaxy keyed 10.7287/peerj.prep
03:36:18 ★★/
■ Abstract ■ Copy

✔ **Transcriptomic analysis reveals how a lack of potassium ions increases Sulfo sensitivity to pH changes**
Microbiology, Vol. 162, No. 8. (01 August 2016), pp. 1422-1434, doi:10.1099/mic.0.000314
by Antoine Buetti-Dinh, Ran Friedman, Olga Dethlefsen, Mark Dopson
posted to methods by galaxyproject to the group Galaxy keyed BuettiDinh2016Transcriptor
■ Copy ■ My Copy

✔ **TP53 gene mutation analysis in chronic lymphocytic leukemia by nanopore MinION sequencing**

Group Tags

All tags in the group Galaxy
Filter:
[Display as Cloud]

methods	1864
workbench	1030
usemain	397
usepublic	373
tools	258
isgalaxy	194
uselocal	184
refpublic	164
cloud	144
other	121
shared	105
reproducibility	97
unknown	72
howto	65
project	54
visualization	27
usecloud	7

Now almost 4000 papers

http://bit.ly/gxycul

The Galaxy Team


Enis Afgan


Dannon Baker


Dan Blankenberg


Dave Bouvier


Marten Cech


John Chilton


Dave Clements


Nate Coraor


Jeremy Goecks


Sam Guerler


Mo Heydarian


Jen Jackson


Vahid Jalili


Delphine Lariviere


Ross Lazarus


Anton Nekrutenko


Nick Stoler


James Taylor

http://wiki.galaxyproject.org/GalaxyTeam

Acknowledgements

You
Karen Eilbeck

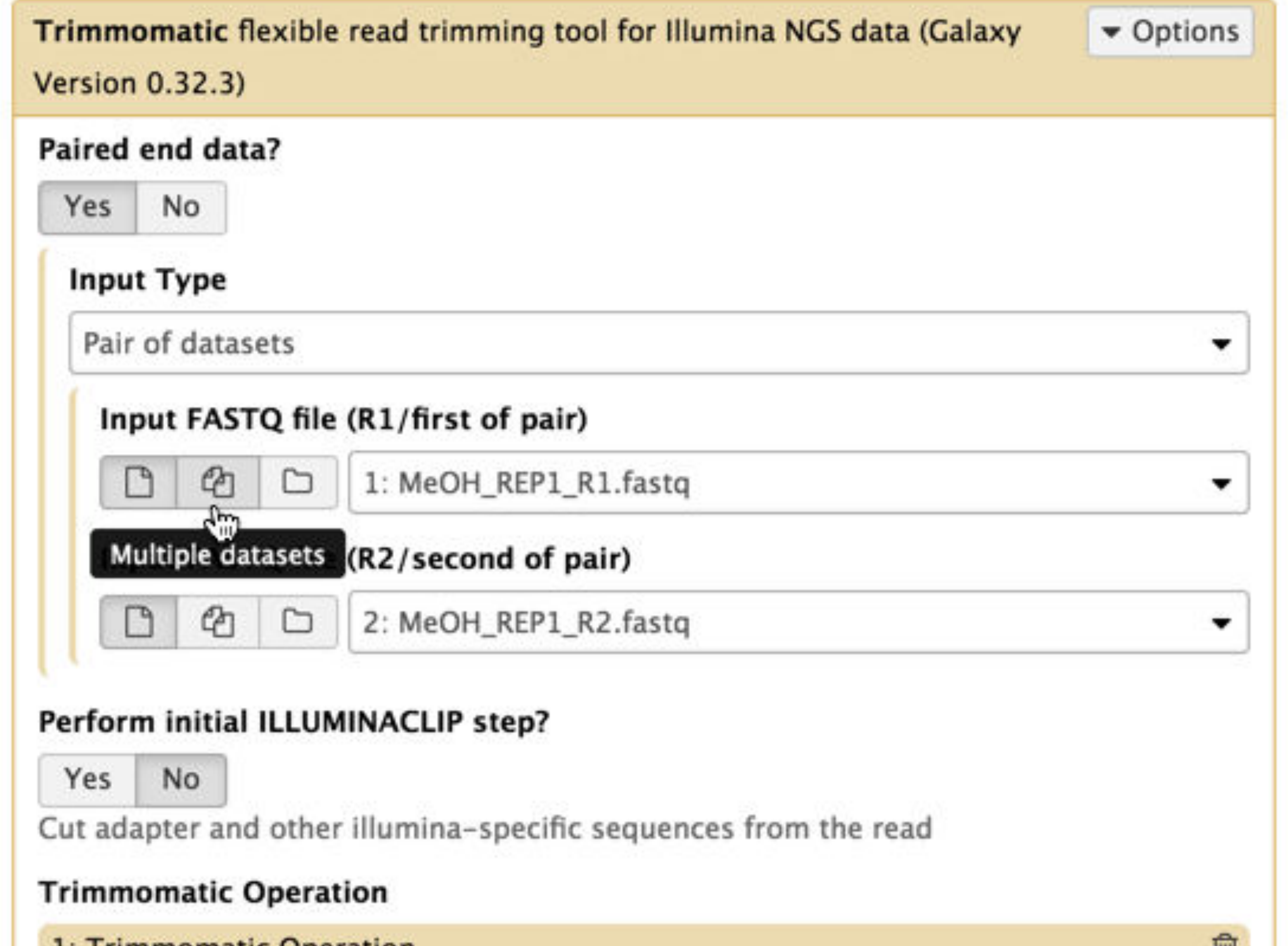
Department of Biomedical Informatics
University of Utah

NIH
XSEDE & Jetstream
Johns Hopkins University
Penn State University

Thanks

And after some resizing and scrolling you see this

Your Friend: The Multiple datasets button



Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.32.3)

Paired end data?

Input Type

Pair of datasets

Input FASTQ file (R1/first of pair)

1: MeOH_REP1_R1.fastq

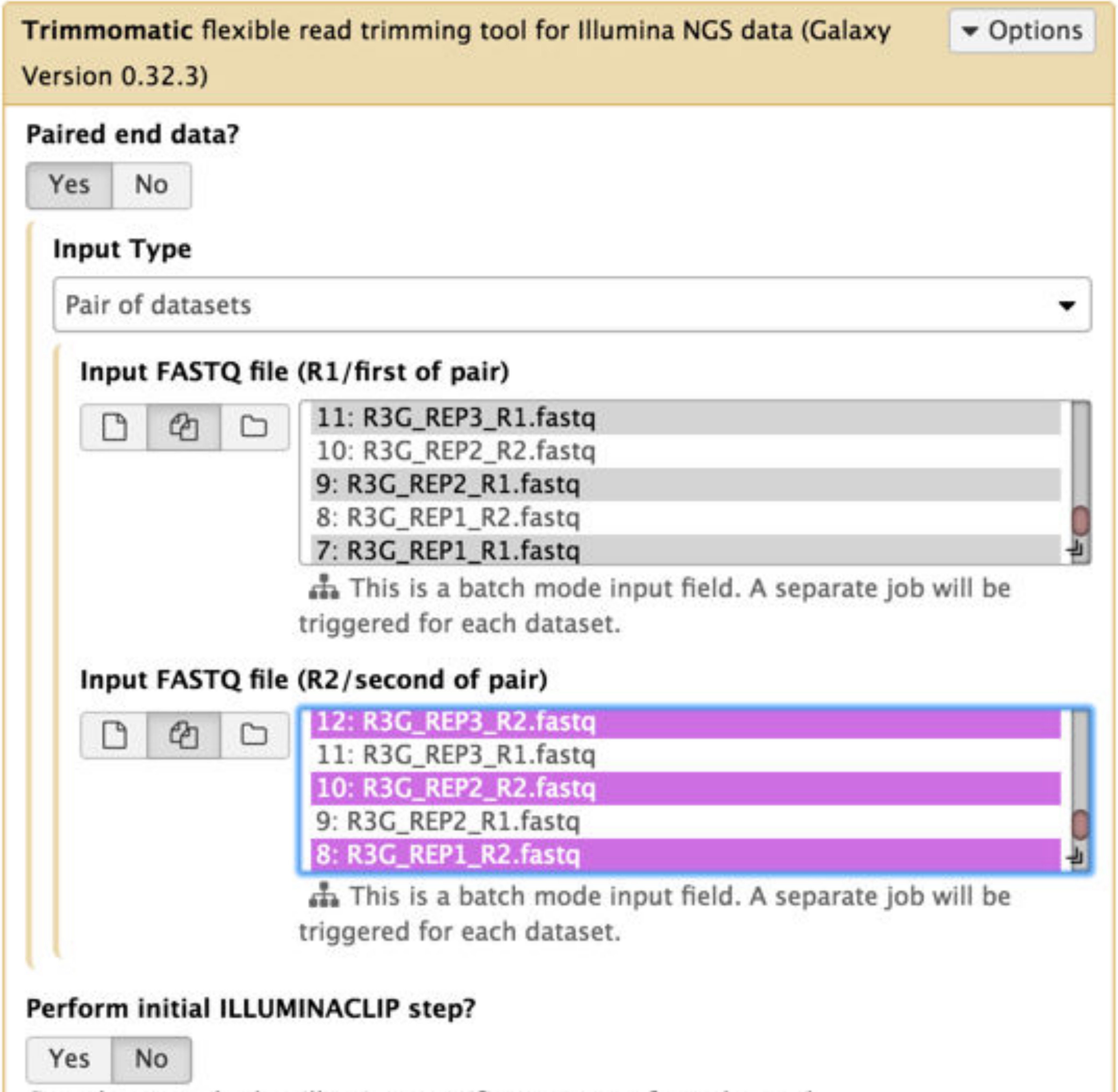
(R2/second of pair)

2: MeOH_REP1_R2.fastq

Perform initial ILLUMINACLIP step?

Cut adapter and other illumina-specific sequences from the read

Trimmomatic Operation



Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.32.3)

Paired end data?

[✓ Yes] No

Input Type

Pair of datasets ▾

Input FASTQ file (R1/first of pair)

- [✓ 11: R3G_REP3_R1.fastq]
- 10: R3G_REP2_R2.fastq
- [✓ 9: R3G_REP2_R1.fastq]
- 8: R3G_REP1_R2.fastq
- [✓ 7: R3G_REP1_R1.fastq]

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

Input FASTQ file (R2/second of pair)

- [✓ 12: R3G_REP3_R2.fastq]
- 11: R3G_REP3_R1.fastq
- [✓ 10: R3G_REP2_R2.fastq]
- 9: R3G_REP2_R1.fastq
- [✓ 8: R3G_REP1_R2.fastq]

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

Perform initial ILLUMINACLIP step?

Yes [✓ No]