



THE  
GERMAN  
HUMAN  
GENOME-  
PHENOME  
ARCHIVE



# Beginner's Guide to Galaxy

Florian Heyl

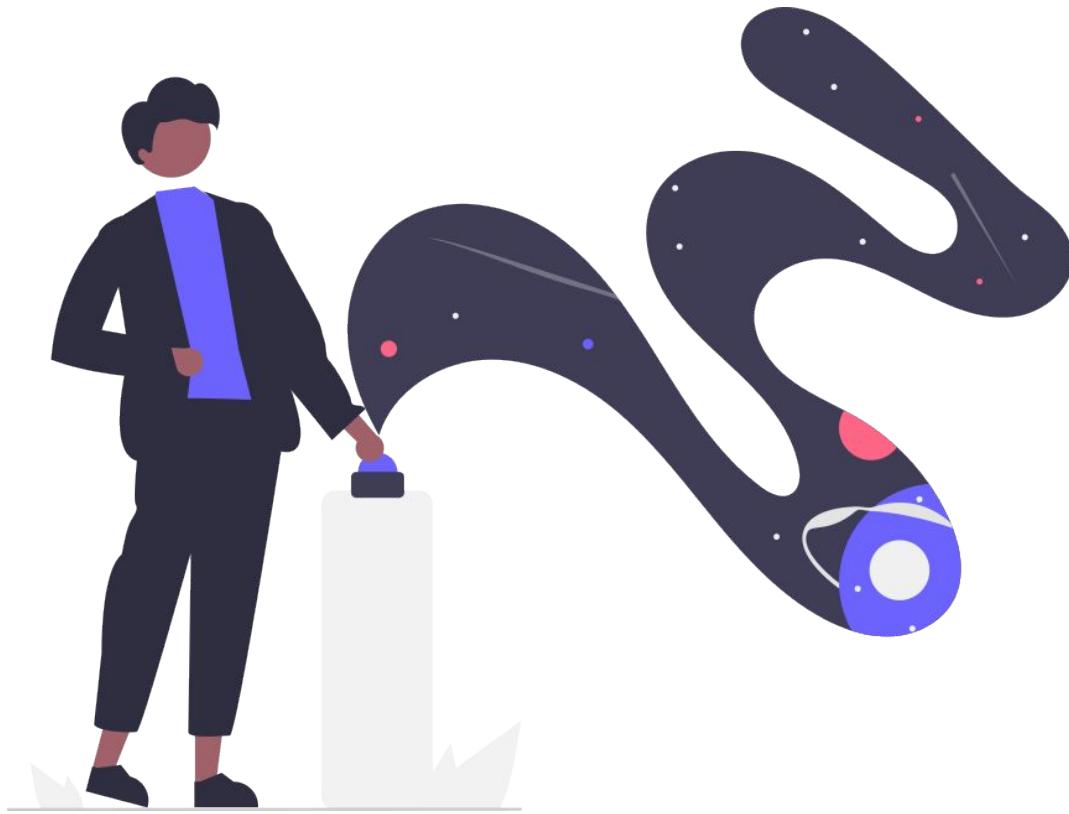
In cooperation with



# Questions to answer

- What is Galaxy? What are the benefits of Galaxy?
- How is Galaxy used? How to start using Galaxy?
- How can you join the Galaxy community?

# What is Galaxy?

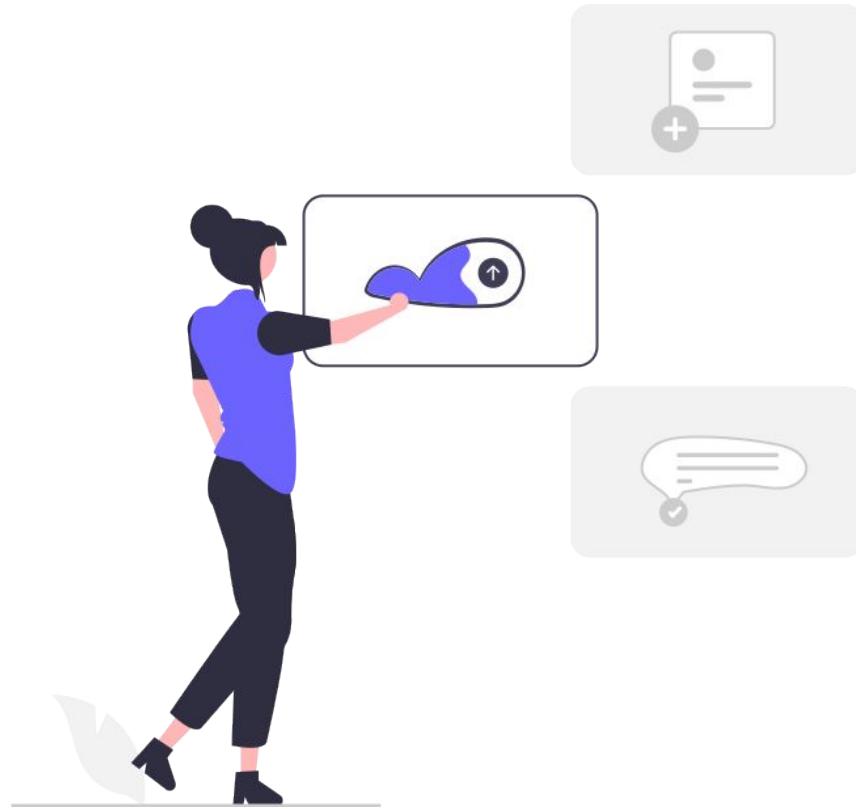


# Data intensive analysis for everyone

- An **open-source** platform for **FAIR** data analysis
- Versatile and reproducible workflows

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar navigation menu includes sections like Tools, Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS (selected), GENOMIC FILE MANIPULATION (selected), COMMON GENOMICS TOOLS, and GENOMICS ANALYSIS. The main workspace displays a "FastQC Read Quality reports (Galaxy Version 0.73+galaxy0)" tool. It has several input fields: "Raw read data from your current history" containing "62: RNA STAR on collection 46: mapped.bam" with a note about batch mode; "Contaminant list" and "Adapter list" both showing "Nothing selected"; "Submodule and Limit specifying file" with "Nothing selected"; and "Lower limit on the length of the sequence to be shown in the report" set to 7. There's also a note about disabling grouping for reads >50bp and a "Length of Kmer to look for" slider set to 7. The right side features a "History" panel listing various analysis steps such as "Sort on collection 85", "MultiQC on data 92", and "featureCounts on collection 62". A top navigation bar includes Workflow, Visualize, Shared Data, Admin, Help, User, and a bell icon.

# What are the benefits of Galaxy?



# 1) Findable



# 1) Findable

usegalaxy.\*: the big 3



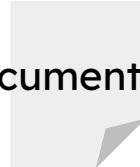
## 2) Accessibility

What did you get in the big 3 usegalaxy.\*?

- Free registration



- Thousands of documented and maintained tools



- 250 GB per user



- Computational resources



# 2) Accessibility

A graphical web interface for running tools interactively

The screenshot shows the Galaxy Europe web interface. The top navigation bar includes links for Workflow, Visualize, Shared Data, Admin, Help, User, and a search bar. A message at the top right indicates "Using 4.1 TB".

The main content area displays the configuration for the "FastQC Read Quality reports (Galaxy Version 0.73+galaxy0)" tool. The configuration includes:

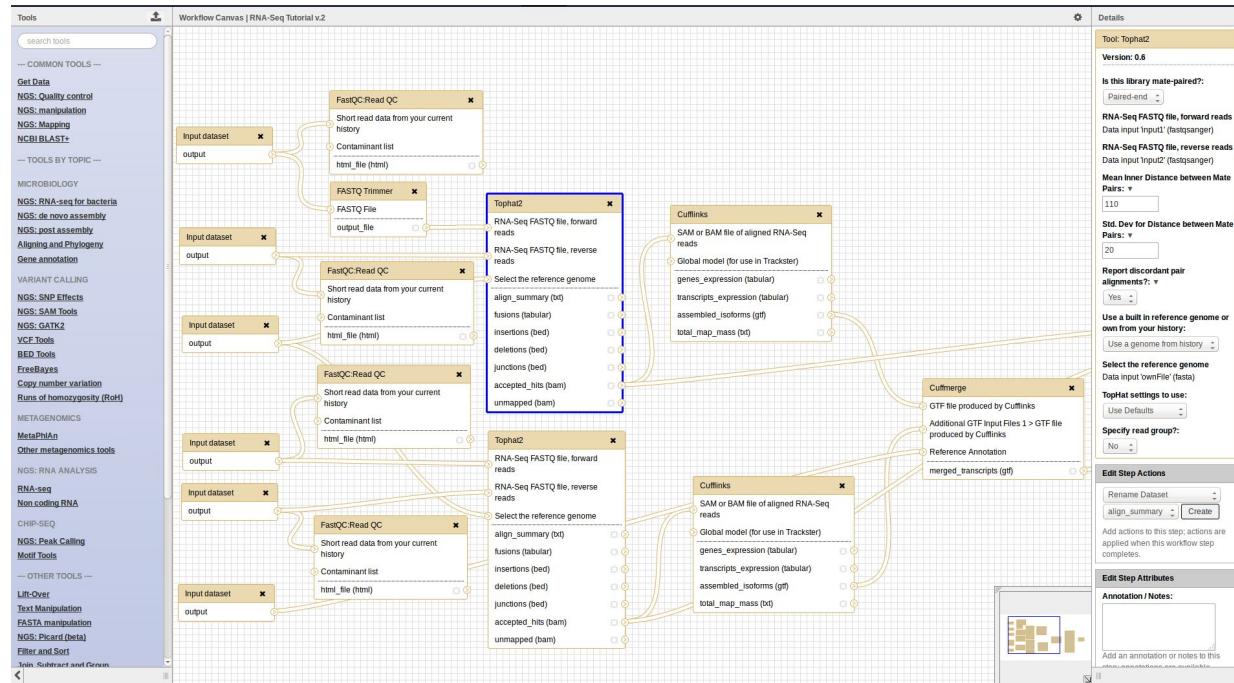
- Raw read data from your current history:** A dropdown menu showing "62: RNA STAR on collection 46: mapped.bam". A note states: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection."
- Contaminant list:** A dropdown menu showing "Nothing selected". A note specifies: "tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAACGGCTACGA"
- Adapter list:** A dropdown menu showing "Nothing selected". A note specifies: "List of adapters adapter sequences which will be explicitly searched against the library. It should be a tab-delimited file with 2 columns: name and sequence. (-adapters)"
- Submodule and Limit specifying file:** A dropdown menu showing "Nothing selected". A note specifies: "a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each submodules warning parameter"
- Disable grouping of bases for reads >50bp:** A radio button set to "No". A note explains: "Using this option will cause fastqc to crash and burn if you use it on really long reads, and your plots may end up a ridiculous size. You have been warned! (-nogroup)"
- Lower limit on the length of the sequence to be shown in the report:** An input field containing "7". A note states: "As long as you set this to a value greater or equal to your longest read length then this will be the sequence length used to create your read groups. This can be useful for making directly comparable statistics from datasets with somewhat variable read lengths. (-min\_length)"
- Length of Kmer to look for:** An input field containing "7". A note states: "Note: the Kmer test is disabled and needs to be enabled using a custom Submodule and limits file (-kmers)"
- Email notification:** A radio button set to "No". A note states: "Send an email notification when the job completes."

At the bottom left is a "Execute" button. On the right side, there is a "History" panel listing various workflow steps:

- 94: Sort on collection 85 (a list with 2 items)
- 95: MultiQC on data 92 a (nd data 89: Webpage)
- 94: MultiQC on data 92 and dat a 89: Stats (a list with 3 items)
- 87: featureCounts on collection 62: Feature lengths (a list with 2 items)
- 86: featureCounts on collection 62: Summary (a list with 2 items)
- 85: featureCounts on collection 62: Counts (a list with 2 items)
- 82: MultiQC on data 80 a (nd data 79: Webpage)
- 81: MultiQC on data 80 and dat a 79: Stats (a list with 2 items)

# 2) Accessibility

A graphical web interface to a powerful workflow system



# 2) Accessibility

>8,000 possible command-line tools

Galaxy Tool Shed

8681 valid tools on Apr 04, 2022

Search

Valid Galaxy Utilities

- Tools
- Custom datatypes
- Repository dependency definitions
- Tool dependency definitions

All Repositories

- Browse by category
- Available Actions
- Login to create a repository

Repositories by Category

Name	Description	Repositories
Assembly	Tools for working with assemblies	181
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	76
Climate Analysis	Tools for analyzing climate data	12
Combinatorial Selections	Tools for combinatorial selection	9
Computational chemistry	Tools for use in computational chemistry	172
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11
Convert Formats	Tools for converting data formats	132
Data Export	Tools for exporting data to various destinations	16
Data Managers	Utilities for Managing Galaxy's built-in data cache	94
Data Source	Tools for retrieving data from external data sources	104
Ecology	Tools related to ecological studies	55
Entomology	Tools that involve insect studies	4
Epigenetics	Tools for analyzing Epigenetic/Epigenomic datasets	35
Fasta Manipulation	Tools for manipulating fasta data	118
Fastq Manipulation	Tools for manipulating fastq data	102
Flow Cytometry Analysis	Tools for manipulating and analyzing FCS files	45
Genome annotation	Tools for annotating genomic information	67
Genome editing	Tools for analyzing genome editing data	9
Genome-Wide Association Study	Utilities to support Genome-wide association studies	24

## 2) Accessibility

Many ways to get data into Galaxy

- Using **Upload Data**
  - Import from your computer
  - Create file by directly entering text
  - Import from a URL
  - Import from S3, DropBox, GoogleDrive
- Using **external public sources** e.g. UCSC, SRA
- From the **shared Galaxy Data Library**

# 3) Interoperable

Share any history and workflow with any user via a link published in the server's shared histories

The screenshot shows the Galaxy Europe web interface. On the left, there is a sidebar with tool categories: Tools, Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS (which is currently selected), and GENOMIC FILE MANIPULATION. The GENERAL TEXT TOOLS section contains sub-options: Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group. The GENOMIC FILE MANIPULATION section contains sub-options: Convert Formats and FASTA/FASTQ.

The main content area has a blue header bar with the Galaxy logo, navigation links (Workflow, Visualize, Shared Data, Admin, Help, User), and a search bar. Below the header, the title "Share or Publish History 'Ref-based RNA-seq - 09.05.22'" is displayed. There are two radio buttons: "Make History accessible" (selected) and "Make History publicly available in Published Histories". A note states: "This History is currently accessible via link and published. Anyone can view and Import this History by visiting the following URL:". Below this is a URL field with the value "url: https://usegalaxy.eu/u/berenice/h/ref-based-rna-seq---training---090522". A modal window titled "Share History with Individual Users" shows a single email address "berenice.batut@gmail.com" entered into a text input field. At the bottom of the modal are "Cancel" and "Save" buttons.

To the right of the main content area, a "History" panel is visible, showing a list of histories. The first history is "Ref-based RNA-seq - 09.05.22" (29 shown, 52 hidden, 10.34 GB). Below it is a list of items:

- 79: Sort on collection 65 (a list with 2 items)
- 75: MultiQC on data 72 and data 69: Webpage
- 74: MultiQC on data 72 and data 69: Stats (a list with 3 items)
- 67: featureCounts on collection 46: Feature lengths (a list with 2 items)

# 4) Reusability / Reproducibility

Detailed **metadata** about each step in an analysis

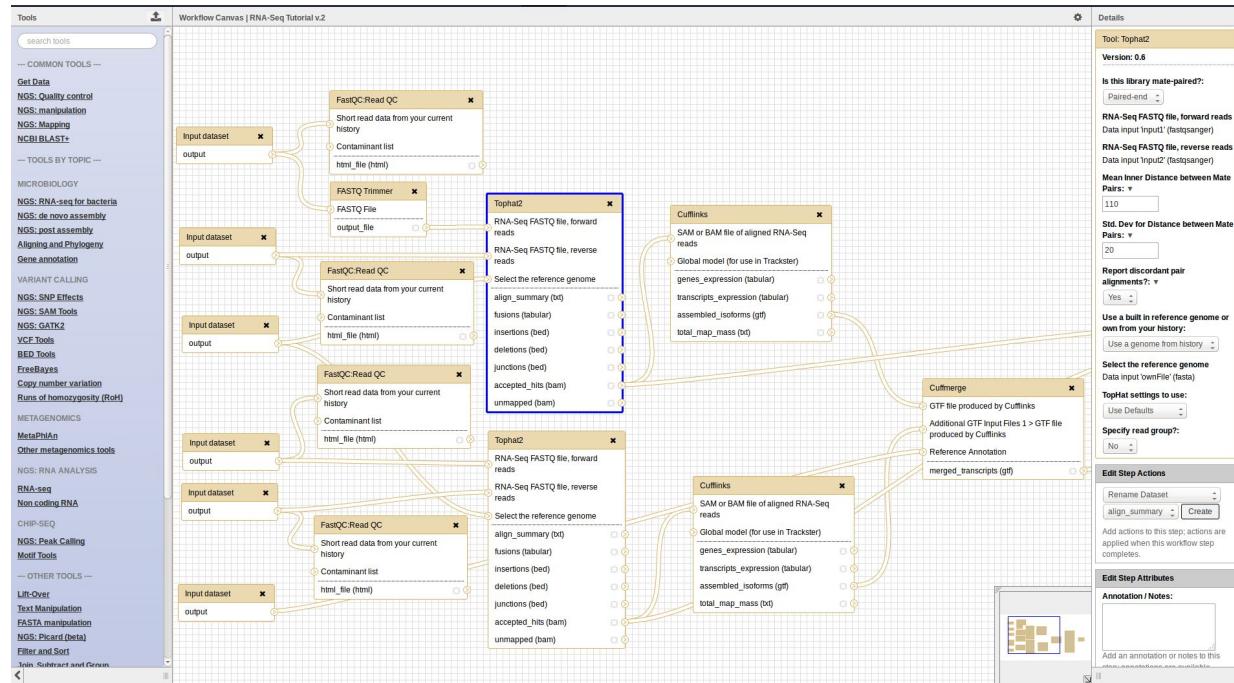
The screenshot shows the Galaxy Europe web interface with the following details:

- Dataset Information:**
  - Number: 81
  - Name: Sort on data 71
  - Created: Tuesday May 10th 8:30:30 2022 UTC
  - Filesize: 255.2 KB
  - Dbkey: dm6
  - Format: tabular
  - File contents: 4838ba20a6d967650755f4fd026a2dab (107275858)
  - History Content API ID: 039d17e8a851e1e (100866)
  - History API ID: 2772b344-78bd-41a2-b76f-f15cb20a7362
  - UUID: 2772b344-78bd-41a2-b76f-f15cb20a7362
  - Full Path: /data/dnb06/galaxy\_db/files/2/7/7/dataset\_2772b344-78bd-41a2-b76f-f15cb20a7362.dat
- Tool Parameters:**
  - Input Parameter: Value
  - Sort Query: 71 featureCounts on data 43 and data 54: Counts (Hidden)
  - Number of header lines: 1
  - on column: 2
  - in: Descending order
  - Flavor: Fast numeric sort (-n)
  - Output unique values: False
  - Ignore case: False
- Job Outputs:**
  - Tool Outputs: Dataset
    - 81 Sort on data 71 (Hidden)
    - 79 Sort on collection 65
    - list of 2 datasets
- Job Information:**
  - Galaxy Tool ID: toolshed.g2.bx.psu.edu/repos/bgruening/text\_processing/tp\_sort\_header\_tool/1.1.1
  - Command Line:

```
( LC_ALL=C sed -u '1'q & sort --stable -t ' ' -k '2rn,2' ) < '/data/dnb06/galaxy_db/files/e/s/3/dataset_ea10f113-419.
```
  - Tool Standard Output: empty
  - Tool Standard Error: empty

# 4) Reusability / Reproducibility

## Workflows!



# 5) Extensive learning resources

[training.galaxyproject.org](https://training.galaxyproject.org)

Galaxy Training!

Contributors Languages Help Extras Search Tutorials

## Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

### Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy Analyses	9
Assembly	13
Climate	6
Computational chemistry	8
Ecology	8
Epigenetics	7
Genome Annotation	10
Imaging	4
Metabolomics	6
Metagenomics	7
Proteomics	26
Sequence analysis	3
Statistics and machine learning	16

### Welcome to the GTN!

Find out more about Galaxy Training Network



Video created by Geert Bonamie.

### The latest GTN news

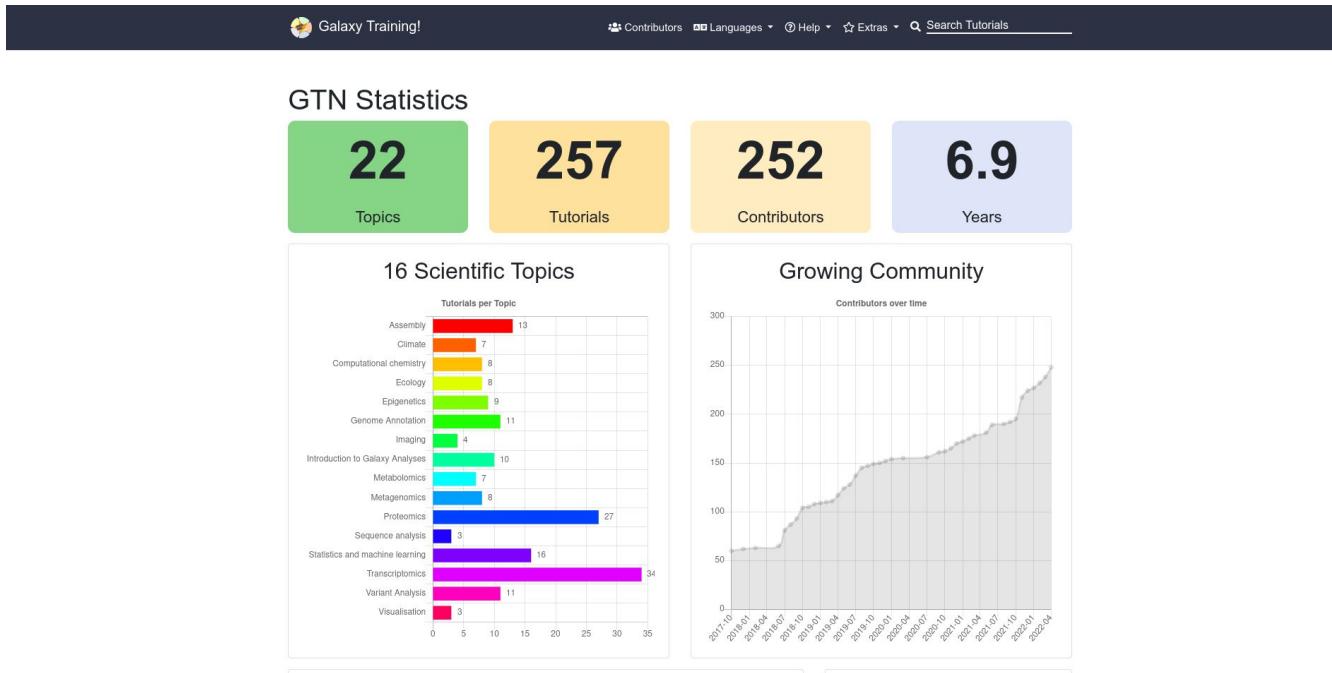
Read about new tutorials, features, events and more!

Apr 12, 2022	New Tutorial Feature: Choose Your Own Tutorial
Mar 14, 2022	New Tutorial: VGP assembly pipeline
Jan 28, 2022	

OPEN CHAT

# 5) Extensive learning resources

[training.galaxyproject.org](https://training.galaxyproject.org)



# How is Galaxy used?

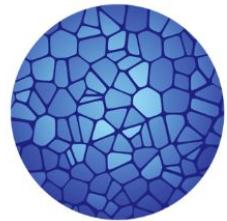


# For which type of data analysis is Galaxy used?

Any data analysis!!!

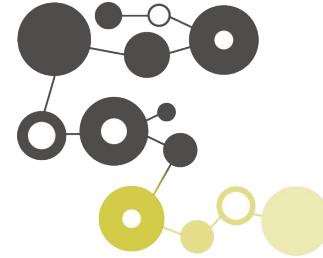
- Transcriptomics
- Epigenetics
- Proteomics
- Metabolomics
- Drug Discovery
- Environment
- Climate Change
- Flow Cytometry
- Natural Language
- Cosmology
- Image Analysis
- Machine learning
- ...

# Single-cell omics



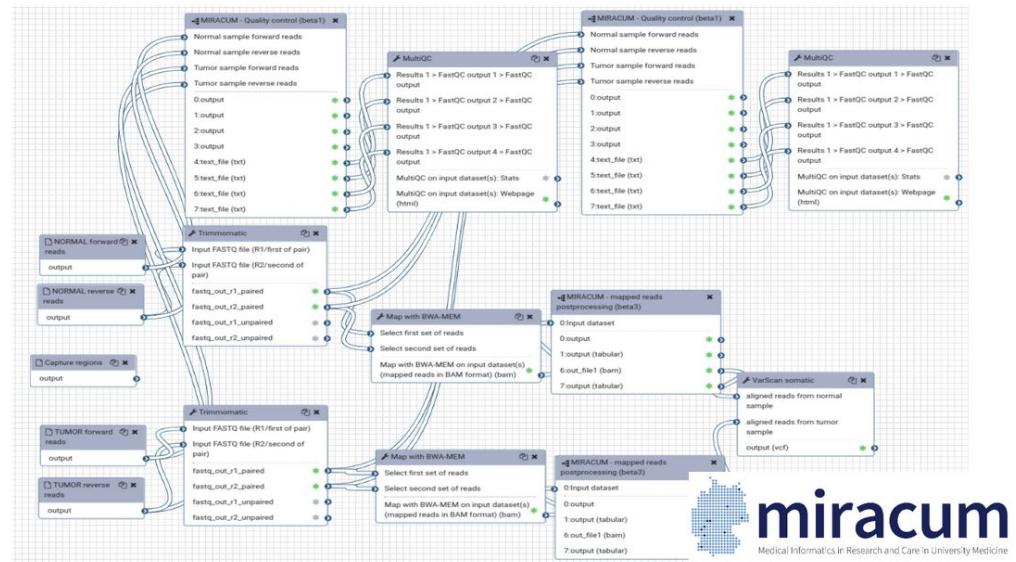
HUMAN  
CELL  
ATLAS

[humancellatlas.usegalaxy.eu](https://humancellatlas.usegalaxy.eu)



[singlecell.usegalaxy.eu](https://singlecell.usegalaxy.eu)

# Somatic Variant Calling and Annotation for informing Molecular Tumor Boards



# Continuous analysis of intra-host variation in SARS-CoV-2

[galaxyproject.org/projects/covid19](https://galaxyproject.org/projects/covid19)

Link	Workflow	Inputs	Outputs	Aligner	Caller
WorkFlowHub DockStore	<b>Illumina ARTIC:</b> Variant analysis from ampliconic data produced with ARTIC protocol v1, v2, v3, or v4, or any alternative primer scheme. <b>ILL-AMP</b>	1. Paired reads [fastq_sanger] 2. SARS-CoV-2 reference [fasta] 3. Primer coordinates [bed] 4. Primer pairs table [tsv]	Variants [vcf]	BWA MEM	lofreq
WorkFlowHub DockStore	<b>Oxford Nanopore ARTIC:</b> Variant analysis from ampliconic data produced with ARTIC protocol v1, v2, v3, or v4, or any alternative primer scheme. <b>ONT-AMP</b>	1. Reads [fastq_sanger] 2. SARS-CoV-2 reference [fasta] 3. Primer coordinates [bed]	Variants [vcf]	minimap2	medaka
WorkFlowHub DockStore	<b>Illumina metatranscriptomic PE:</b> Variant analysis from metatranscriptomic data. <b>ILL-MT-PE</b>	1. Paired reads [fastq_sanger] 2. SARS-CoV-2 reference [fasta]	Variants [vcf]	BWA MEM	lofreq
WorkFlowHub DockStore	<b>Illumina metatranscriptomic SE:</b> Variant analysis from metatranscriptomic data. <b>ILL-MT-SE</b>	1. Reads [fastq_sanger] 2. SARS-CoV-2 reference [fasta]	Variants [vcf]	BWA MEM	lofreq
WorkFlowHub DockStore	<b>Report generation:</b> Generation of final variant analysis reports/plots. <b>REPORTING</b>	1. Variants [vcf] 2. Gene name translation table [tsv]	Reports [tsv], overview [svg]	-	-

vcf = variant call format, tsv = TAB-separated values, svg = scalable vector graphics, fastq\_sanger = fastq format with Sanger encoding of base quality values, bed = browser extensible format

# Check our Galaxy Publication Library

[Log In · Register](#)

[Upgrade Storage](#)

[Home](#) [Groups](#) [Documentation](#) [Forums](#) [Get Involved](#)

[Search for groups](#)  [Search](#)

[Home > Groups > Galaxy](#)

## Galaxy

Group Library

### Recently Added Items

Title	Added By	Date Modified
<a href="#">MaxQuant and MSstats in Galaxy Enable Reproducible Cloud-Bases...</a>	bag	5/4/2022, 20:34:53
<a href="#">Galaxy workflows for fragment-based virtual screening: a case...</a>	bag	4/14/2022, 13:38:05
<a href="#">Ten simple rules for making a software tool workflow-ready</a>	bag	4/10/2022, 17:34:38
<a href="#">Small Conductance Ca<sup>2+</sup>-Activated K<sup>+</sup> (SK) Channel mRNA Expression</a>	bag	3/19/2022, 00:08:26
<a href="#">Gut microbiota drives age-related oxidative stress and mitoc...</a>	bag	3/7/2022, 10:18:22
<a href="#">A <i>Tetragenococcus halophilus</i> human gut isolate</a>	bag	2/21/2022, 15:24:50
<a href="#">ZC3H4 restricts non-coding transcription in human cells</a>	bag	2/21/2022, 14:22:28
<a href="#">Within-host evolution of SARS-CoV-2 in an immunosuppressed C...</a>	bag	2/21/2022, 14:22:28
<a href="#">VPRP functions downstream of the androgen receptor and OGT...</a>	bag	2/21/2022, 14:22:28
<a href="#">Uncovering the RNA-binding protein landscape in the pluripot...</a>	bag	2/21/2022, 14:22:28

See all 11827 items for this group in the [Group Library](#).

## Galaxy

Publication library for the [Galaxy Project](#). Contains publications that use, reference, implement, and extend Galaxy.

Galaxy is an open, web based platform for data integration and analysis, primarily in the life sciences.

Group [tags are explained here](#).

Want to contribute? [Here's how](#).

<https://galaxyproject.org/>

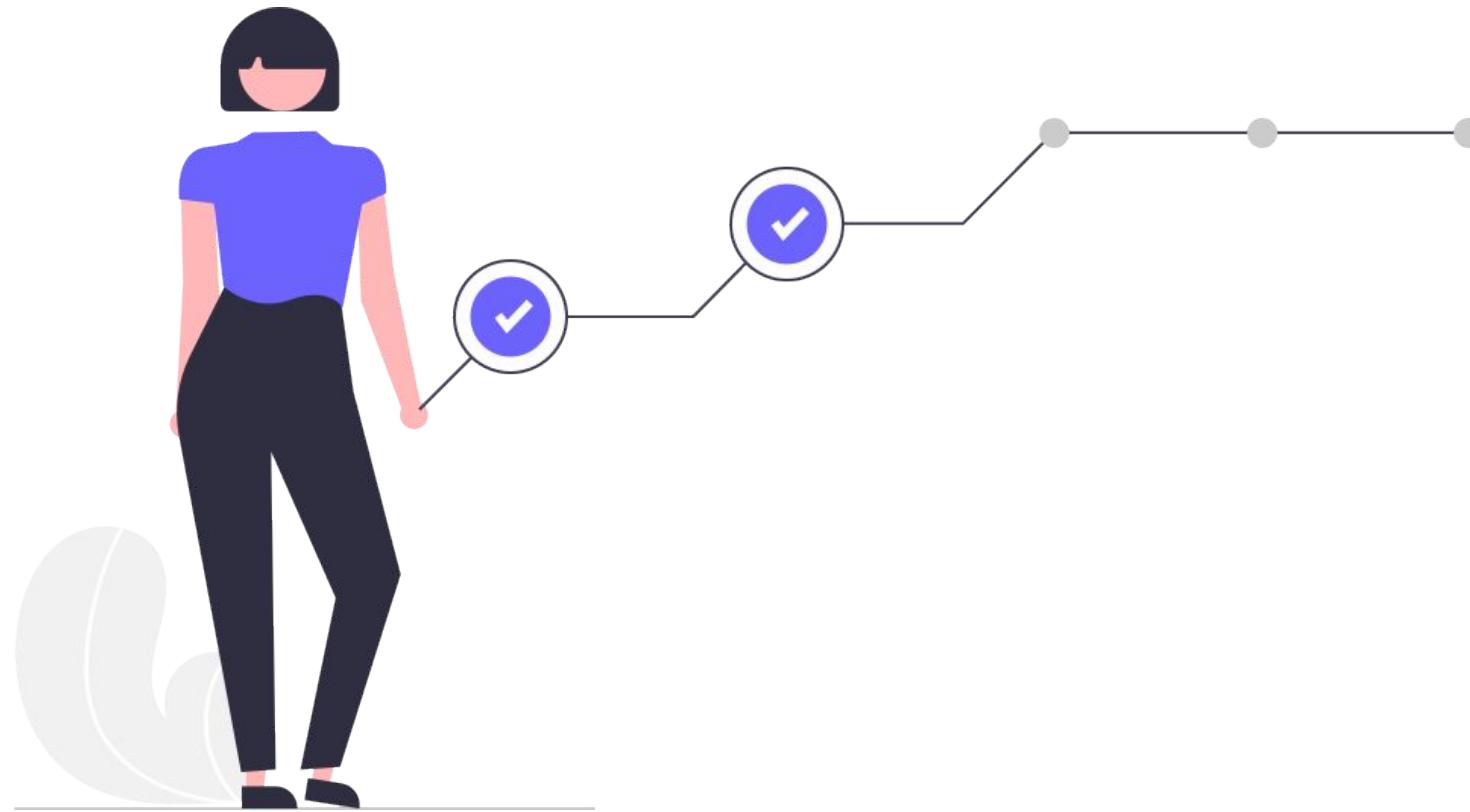
Owner: [Galaxy Project](#)  
Registered: 2017-09-18  
Type: Public  
Membership: Closed  
Library Access: You can only view

[Log in](#) or [Register](#) to join groups

### Members (4)



# How to start using Galaxy?



# Galaxy interface - Let's get into it.

The screenshot displays the Galaxy web interface on a Mac OS X system. The window is divided into three main sections:

- Left Panel (Blue Border): Tools Catalog**
  - Search bar: "search tools"
  - Upload Data button
  - Get Data
  - Collection Operations
  - GENERAL TEXT TOOLS** (highlighted)
  - Text Manipulation
  - Filter and Sort
  - Join, Subtract and Group
  - Datamash
  - GENOMIC FILE MANIPULATION**
  - FASTA/FASTQ
  - FASTQ Quality Control
  - SAM/BAM
  - BED
  - VCF/BCF
  - Nanopore
  - Convert Formats
  - Lift-Over
  - COMMON GENOMICS TOOLS**
  - Interactive tools
  - Operate on Genomic Intervals
  - Fetch Sequences/Alignments
  - GENOMICS ANALYSIS**
  - Assembly
- Middle Panel (Red Border): Home Page**

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.

**James P. Taylor Foundation for Open Science.**

"The most important job of senior faculty is to mentor junior faculty and students." — @jptx

Text Text

Announcing the James P. Taylor (JXTX) Foundation for Open Science

[Learn More](#)

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the Galaxy SARS-CoV-2 portal at [covid19.galaxyproject.org](#)

**PennState** **JOHNS HOPKINS UNIVERSITY** **OREGON HEALTH & SCIENCE UNIVERSITY**

The Galaxy Team is a part of the Center for Comparative Genomics and Evolutionary Biology at the University of Illinois Urbana-Champaign.
- Right Panel (Green Border): History Panel**

History

search datasets

**Galaxy 101 History**

2 shown 7.48 MB

2: SNPs

1: Exons

# Top menu



Link	Usage
(or <i>Analyze Data</i> )	go back to the homepage
<i>Workflow</i>	access existing workflows or create new one using the editable diagrammatic pipeline
<i>Visualize</i>	create new visualisations and launch Interactive Environments
<i>Shared Data</i>	access data libraries, histories, workflows, visualizations and pages shared with you
<i>Help</i>	links to Galaxy Help Forum (Q&A), Galaxy Community Hub (Wiki), and Interactive Tours
<i>User</i>	your preferences and saved histories, datasets, pages and visualizations

# Tools

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools

join the intervals of two datasets side-by-side (Galaxy Version 1.0.0)

**Join**

1: Exons

First dataset

with

2: SNPs

Second dataset

with min overlap

1  
(bp)

Return

Only records that are joined (INNER JOIN)

Execute

**TIP:** If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

**Screencasts:**  
See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

**Syntax**

- Where overlap specifies the minimum overlap between intervals that allows them to be joined.
- Return only records that are joined returns only the records of the first dataset that join to a record in the second dataset. This is analogous to an INNER JOIN.
- Return all records of first dataset (fill null with ".") returns all intervals of the first dataset, and any intervals that do not join an interval from the second dataset are filled in with a period(.). This is analogous to a LEFT JOIN.
- Return all records of second dataset (fill null with ".") returns all intervals of the second dataset, and any intervals that do not join an interval from the first dataset are filled in with a period(.). Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.
- Return all records of both datasets (fill nulls with ".") returns all records from both datasets, and fills on either the right or left with periods. Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.

History

search datasets

Galaxy 101

2 shown, 5 deleted

9.06 MB

2: SNPs

1: Exons

# Tools

- A tool form contains:
  - input datasets and parameters
  - help, citations, metadata
  - an Execute button to start a job
- New tool versions can be installed without removing old ones to ensure reproducibility

Sort data in ascending or descending order (Galaxy Version 1.2.0)

Sort Dataset

on column

with flavor

Numerical sort

everything in

Descending order

Column selection

+ Insert Column selection

Number of header lines to skip

0

# characters are already considered as comments and kept

Email notification

No

Send an email notification when the job completes.

Execute

 **TIP:** If your data is not TAB delimited, use *Text Manipulation->Convert*

## Syntax

This tool sorts the dataset on any number of columns in either ascending or descending order.

- Numerical sort orders numbers by their magnitude, ignores all characters besides numbers, and evaluates a string of numbers to the value they signify.

# Free "app" store: Galaxy Tool Shed

- Thousands of tools already available
- Most software can be integrated
- If a tool is not available, ask the Galaxy community for help!
- Only a Galaxy admin can install tools

The screenshot shows the Galaxy Tool Shed homepage. At the top, there's a navigation bar with links for 'Repositories', 'Groups', 'Help', and 'User'. Below the navigation, a message states '6532 valid tools on Dec 04, 2018'. On the left, there's a sidebar with a 'Search' section containing links for 'Search for valid tools', 'Search for workflows', 'Valid Galaxy Utilities' (with links for 'Tools', 'Custom datatypes', 'Repository dependency definitions', and 'Tool dependency definitions'), 'All Repositories' (with a 'Browse by category' link), and 'Available Actions' (with a 'Login to create a repository' link). The main content area is titled 'Repositories by Category' and contains a table with columns for 'Name', 'Description', and 'Repositories'. The table lists several categories: 'Assembly' (Tools for working with assemblies, 128 repositories), 'ChIP-seq' (Tools for analyzing and manipulating ChIP-seq data, 65 repositories), 'Combinatorial Selections' (Tools for combinatorial selection, 10 repositories), 'Computational chemistry' (Tools for use in computational chemistry, 76 repositories), 'Constructive Solid Geometry' (Tools for constructing and analyzing 3-dimensional shapes and their properties, 12 repositories), 'Convert Formats' (Tools for converting data formats, 114 repositories), and 'Tools for exporting data to various' (1 repository).

Name	Description	Repositories
Assembly	Tools for working with assemblies	128
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	65
Combinatorial Selections	Tools for combinatorial selection	10
Computational chemistry	Tools for use in computational chemistry	76
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	12
Convert Formats	Tools for converting data formats	114
	Tools for exporting data to various	~

# History

- Location of all analyses history
  - collects all datasets produced by tools
  - collects all operations performed on the data
- For each dataset (the heart of Galaxy's reproducibility), the history tracks
  - name, format, size, creation time, datatype-specific metadata
  - tool id, version, inputs, parameters
  - standard output (stdout) and error (stderr)
  - state (waiting, running, success, failed)
  - hidden, deleted, purged

The screenshot shows the Galaxy History interface. At the top, there is a search bar labeled "search datasets" and a section titled "Galaxy 101" showing "7 shown" datasets totaling "9.07 MB". Below this, a green header bar indicates "7: Compare two Datasets on data 6 and data 1". The main area displays a BED file named "hg38" with 5 regions. The file is a join operation (join (GNU coreutils) 8.22) using the GNU GPL license. A note states: "This is free software: you are free to change and redistribute it. There is NO WARRANTY; to the extent". Below the file details, there are buttons for "display in IGB View", "display with IGV local Human hg38", and "display at UCSC main test". A table below lists the genomic coordinates:

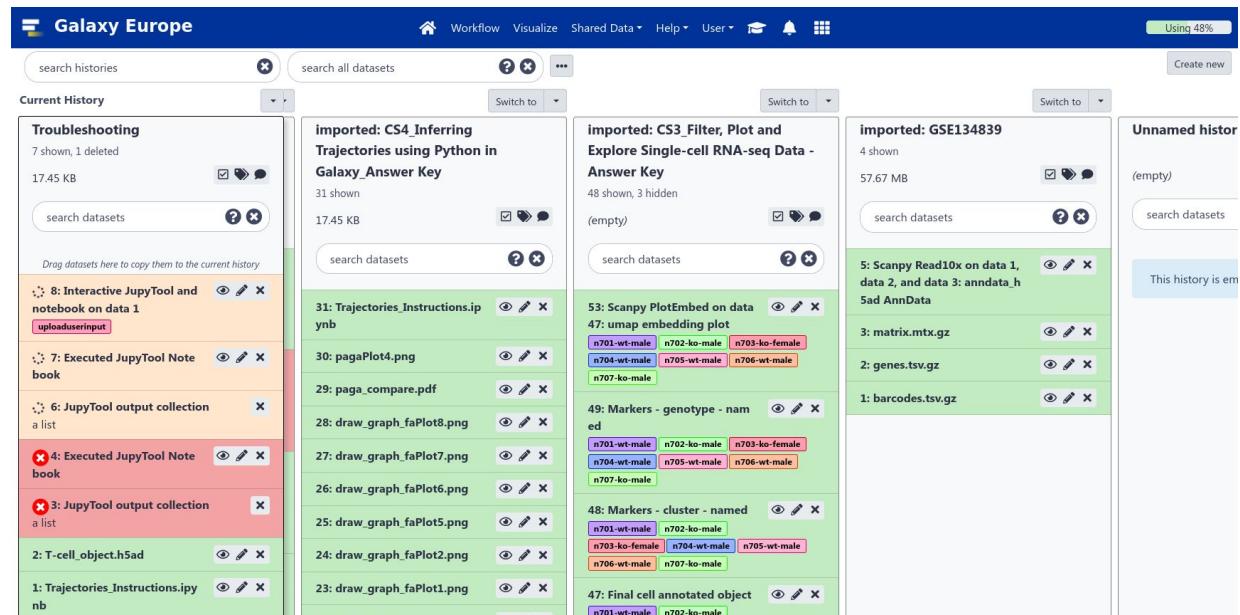
1. Chrom	2. Start	3. End	4. Name
chr22	46256560	46263322	uc003bhh.
chr22	15690077	15690709	uc010gqp.
chr22	15528158	15529139	uc011agd.
chr22	15690245	15690709	uc062bek.
chr22	22376182	22376505	uc062cbs.

Below the table, a series of operations are listed in a stack:

- 6: Select first on data 5
- 5: Sort on data 4
- 4: Group on data 3
- 3: Join on data 2 and data 1
- 2: SNPs

# All histories

- You can have as many histories as you want
  - each history should correspond to a different analysis
  - and should have a meaningful name



# Importing data

- Copy/paste some text
- Upload files from your local computer
- Upload data from an internet URL
- Upload data from online databases: UCSC, BioMart, ENCODE, etc.
- Import from Shared Data (libraries, histories, pages)
- Upload data from FTP

See: [Galaxy tutorial for important data](#)

# Data types

- Tools only accept input datasets with the appropriate datatypes
- When uploading a dataset, its datatype can be either:
  - automatically detected
  - assigned by the user
- Datasets produced by a tool have their datatype assigned by the tool
- To change the datatype of a dataset, either:
  - Edit attributes and Datatypes (if original wrong), or
  - Edit attributes and Convert

# Reference datasets

- Genome build specifies which genome assembly a dataset is associated with
- Can be assigned by a tool or by the user
- Users can create custom genome builds
- New builds can be added by the admin

## Database/Build

Mouse July 2007 (NCBI37/mm9) (mm9)

Burmese python Sep. 2013 (Python\_molurus\_bivittatus-5.0.2/pytBiv1) (pytBiv1)

Burton's mouthbreeder Oct 2011 (AstBur1.0/hapBur1) (hapBur1)

Bushbaby Mar. 2011 (Broad/otoGar3) (otoGar3)

Bushbaby Dec. 2006 (Broad/otoGar1) (otoGar1)

C. angaria Oct. 2010 (WS225/caeAng1) (caeAng1)

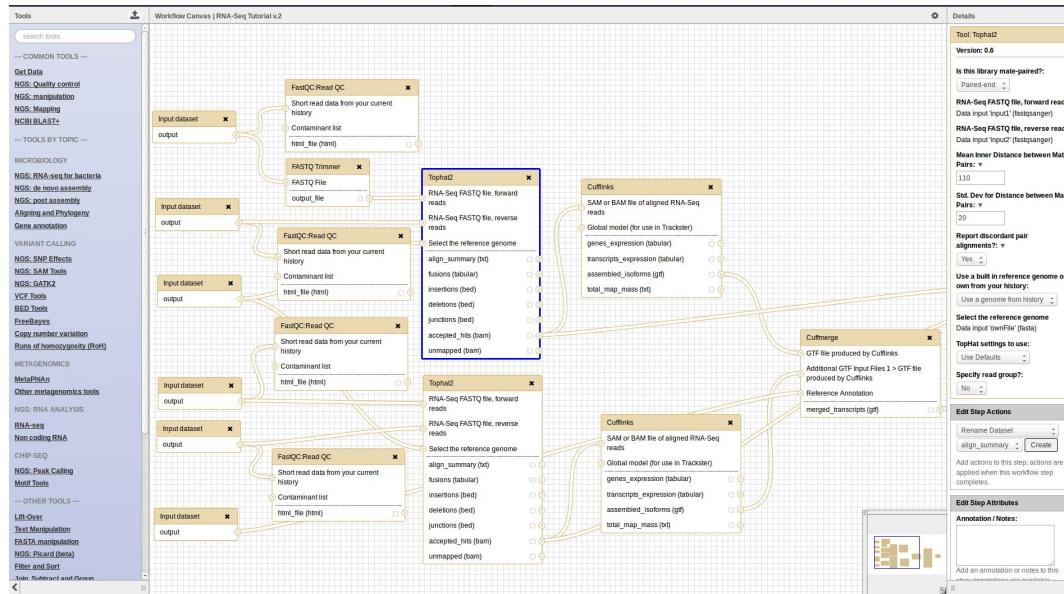
C. brenneri Nov. 2010 (C. brenneri 6.0.1b/caePb3) (caePb3)

C. brenneri Feb. 2008 (WUGSC 6.0.1/caePb2) (caePb2)

C. brenneri Jan. 2007 (WUGSC 4.0/caePb1) (caePb1)

# Workflow Editor

- **Extracted from a history**
- **Built manually** by adding and configuring tools using the canvas
- **Imported** using an existing shared workflow

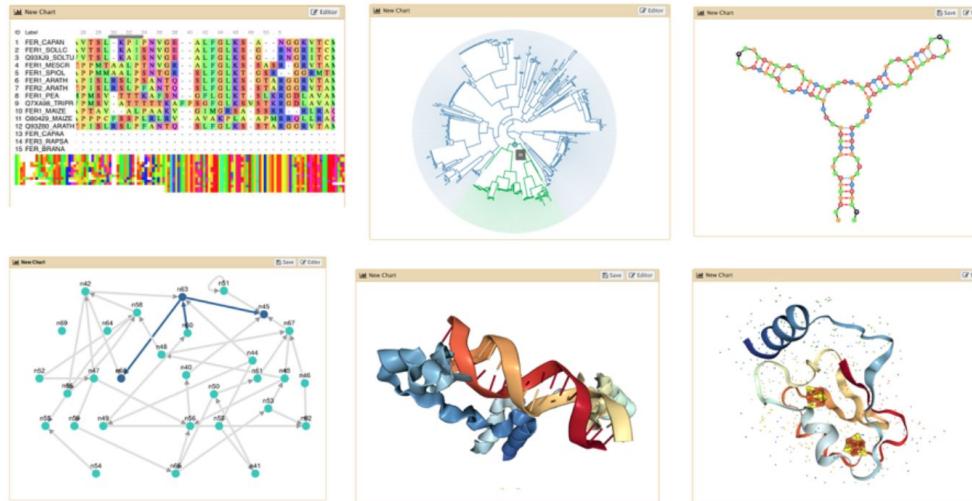


# Why would you want to create workflows?

- **Re-run** the same analysis on different input data sets
- **Change parameters** before re-running a similar analysis
- Make use of the workflow **job scheduling** (jobs are submitted as soon as their inputs are ready)
- Create **sub-workflows**: a workflow inside another workflow
- **Share** workflows for publication and with the community

# Visualizations

- Datatypes know what tools can be used to visualize datasets:
    - Sequencing data has a button for visualizing in IGV
    - Tabular data will prompt you to build charts
    - Protein data can be seen in a 3D viewer
  - Interactive environments: Jupyter, RStudio, etc

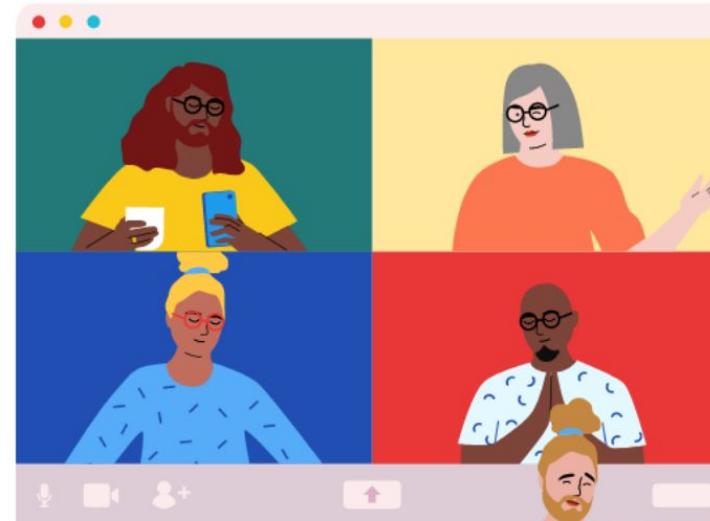


# Sharing

- Share everything you do in Galaxy - histories, workflows, and visualizations
  - Directly using a Galaxy account's email addresses on the same instance
  - Using a web link, with anyone who knows the link
  - Using a web link and publishing it to make it accessible to everyone from the Shared Data menu

See: [Sharing your History in Galaxy](#)

# How can you join the Galaxy community?



# Several channels

- Support forum: [Galaxy Help](#)
- Community curated documentation: [Galaxy Community Hub](#)
- [Events](#) all around the world
- Galaxy Training for scientists, developers, admins, instructors: [Galaxy Training Community](#)
- Training questions? Chat with us on [Gitter](#)

The screenshot shows a forum interface with the title "GalaxyHelp". The top navigation includes a search bar, "Sign Up", "Log In", and a menu icon. Below the header, there are buttons for "all categories", "all tags", and "Latest" (which is highlighted in red). Other buttons include "Top" and "Categories". A table lists forum posts. The first post, by "usegalaxy.org support", has a yellow profile picture and a small icon. It has 1 reply, 85 views, and was posted 7 days ago. The second post, also by "usegalaxy.org support", has a yellow profile picture and a small icon. It has 1 reply, 75 views, and was posted 15 days ago. Both posts are related to troubleshooting errors.

Topic	Category	Users	Replies	Views	Activity
🔒 Troubleshooting resources for errors or unexpected results Start by reviewing the troubleshooting FAQ. Common reasons and solutions for tool errors are explained. Most job errors can be resolved by correcting your input data's format/content. Others indicate a tool setting/param... <a href="#">read more</a>	usegalaxy.org support		1	85	7d
🔒 Welcome to Galaxy Community Help For assistance with a specific Galaxy server please post into appropriate category.	usegalaxy.org support		1	75	15d

# Galaxy Mentor Network



Galaxy Mentor Network

About Guides ▾ FAQ Network ▾ Blog Code of Conduct

Mentorship... for you!

An all in one program targeted to speed up your growth!

Apply Now



Community

Birthed from the Galaxy Community, with it at  
the center.



Collaboration

The program is founded and sustained by the  
pillars of collaboration.



Growth

Your growth and expectations are our focus,  
and your growth drives us.

# Events: Smörgåsbord 2023 (22.05)

*The Gallantries, Galaxy Training Network & Galaxy Community are happy to announce*

# GTN Smörgåsbord 3 22-26 May 2023

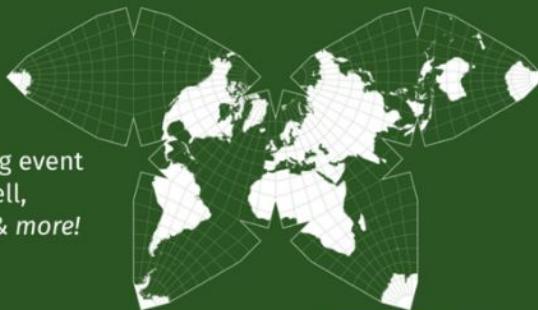
Save the date!

[gxy.io/smorgasbord3](https://gxy.io/smorgasbord3)

Join a **free, global**, week-long Galaxy Training event covering everything from RNA-Seq, Single Cell, Proteomics, SARS-CoV-2, Cancer, RO-Crates & more!



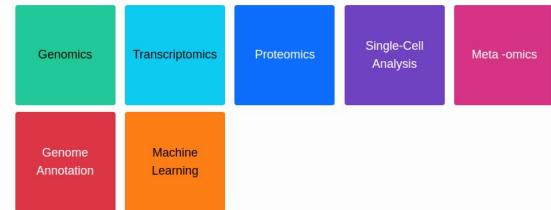
@gtn@mstdn.science  
@gallantries@mstdn.science  
@gxytraining @Gallantries\_EU



With the support  
of the  
European Union

## Browse by Analysis Type

I'm interested in...



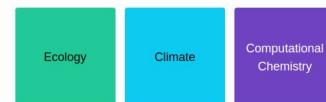
## Browse by Sample Type

I want to analyze...



## Non-Genomics modules

I want to learn about...



## FAIR data & Best Practices



## Events: GCC 2023 (10-16.06, Australia)





# Thanks to the Galaxy community:

- Andrea Bagnacani
- Bérénice Batut
- Saskia Hiltemann
- Anne Pajon
- Nicola Soranzo
- Helena Rasche
- Christopher Barnett
- Michele Maroni
- Anne Fouilloux
- AvatarNadia Goué
- Olha Nahorna
- Dave Clements

Illustrations and content provided  
by: Bérénice Batut



Photo by Bérénice Batut on Flickr