

# Introduction to Galaxy

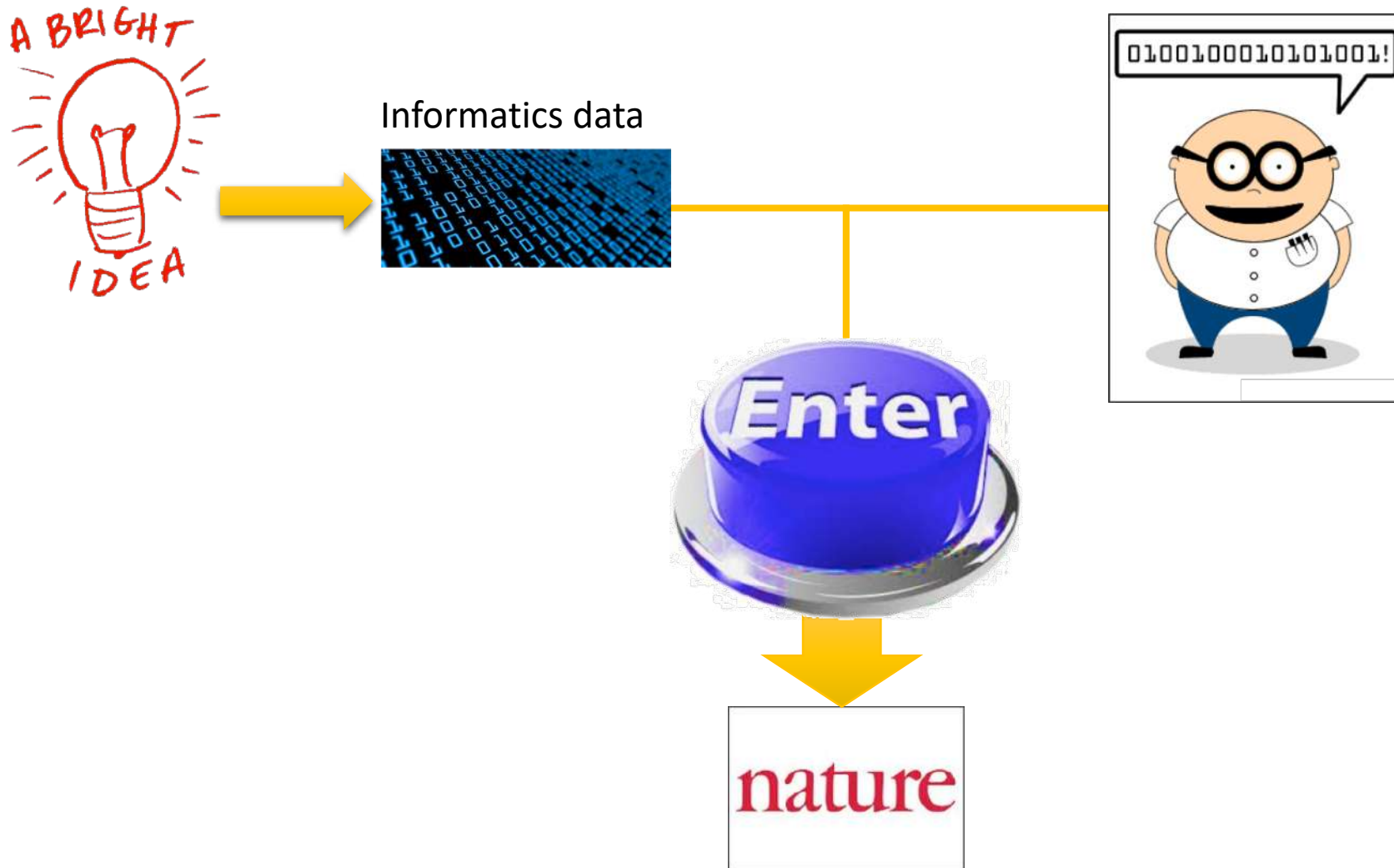
Stéphanie Le Gras  
([slegras@igbmc.fr](mailto:slegras@igbmc.fr))

# Guidelines

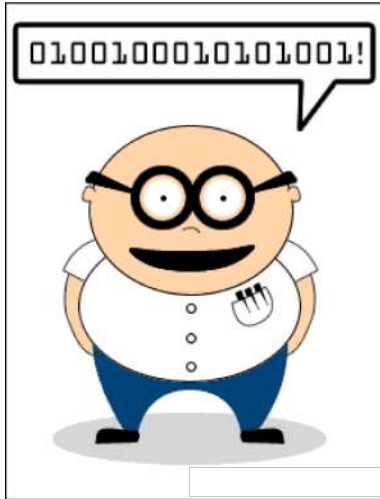
- Analyzing biological data with informatics tools
- Presentation of the Galaxy project
- Description of the main features of the Galaxy platform

# Analyzing biological data with informatics tools

# Bioinformatics analyses



# Bioinformatics analyses



## Scripts, softwares

```
#!/usr/bin/perl

use strict;
use warnings;
use Getopt::Long;

## Date : 22 fev 2011
## Author : Stephanie Le Gras

## Objectives :

my $num_arg = scalar @ARGV;
my $programe = "ExtractID.pl";
my $input;
my $out;
my $id;

my $result = GetOptions(
    "id=s" => \$id,
    "out=s" => \$out,
    "input=s" => \$input,
);

my $usage = <<END;

Usage: $programe --id=FILENAME --out=FILENAME --input=FILENAME

END

die $usage unless ($result);
my @files = @ARGV;
die "Enter at least two files\n$usage" if ( $num_arg < 2 );
die $usage if ( $num_arg == 0 );

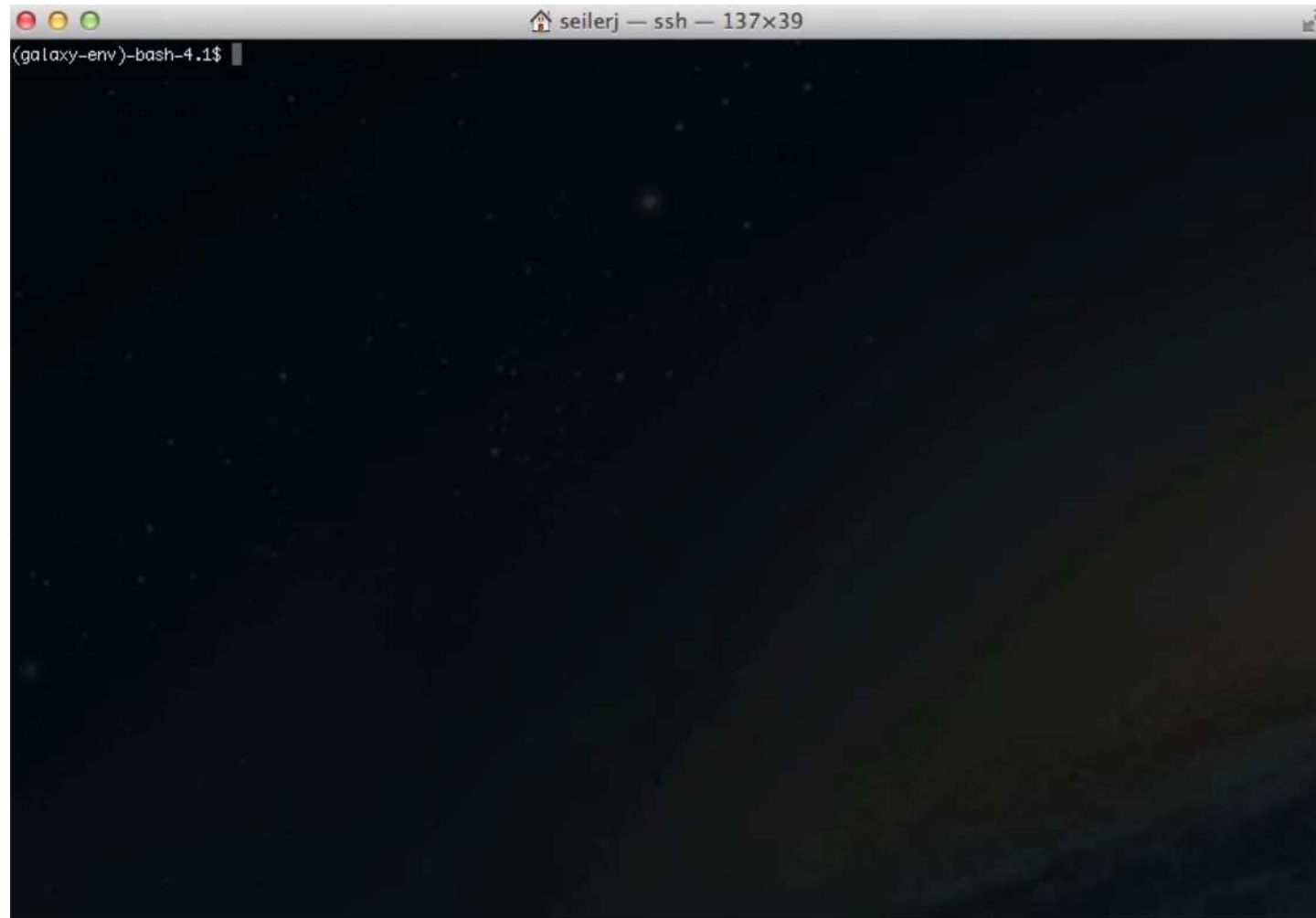
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

## first, every lines of each files are put in the hash table ids. Variant ids are used as keys of the
## hash table and it contains a table.
```

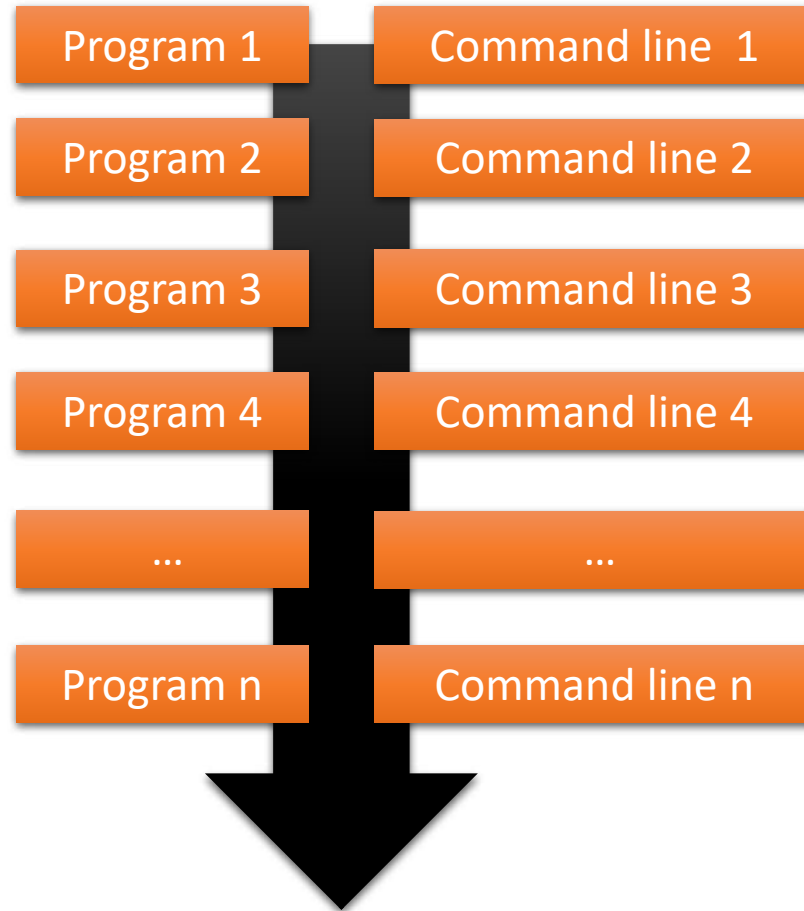
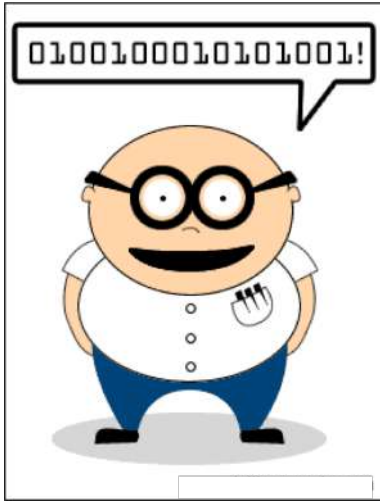
## Command line

```
macs14 -t treatment.sort.bed -c control.bed -f BED -g mm --name=name1 --llocal=50000
--slocal=5000 > macs1.nohup 2>&1 &
```

# Bioinformatics analyses



# Bioinformatics analyses



**PIPELINE /  
WORKFLOW**

# Galaxy ?







# Galaxy

PROJECT

# Galaxy project

# What is Galaxy ?

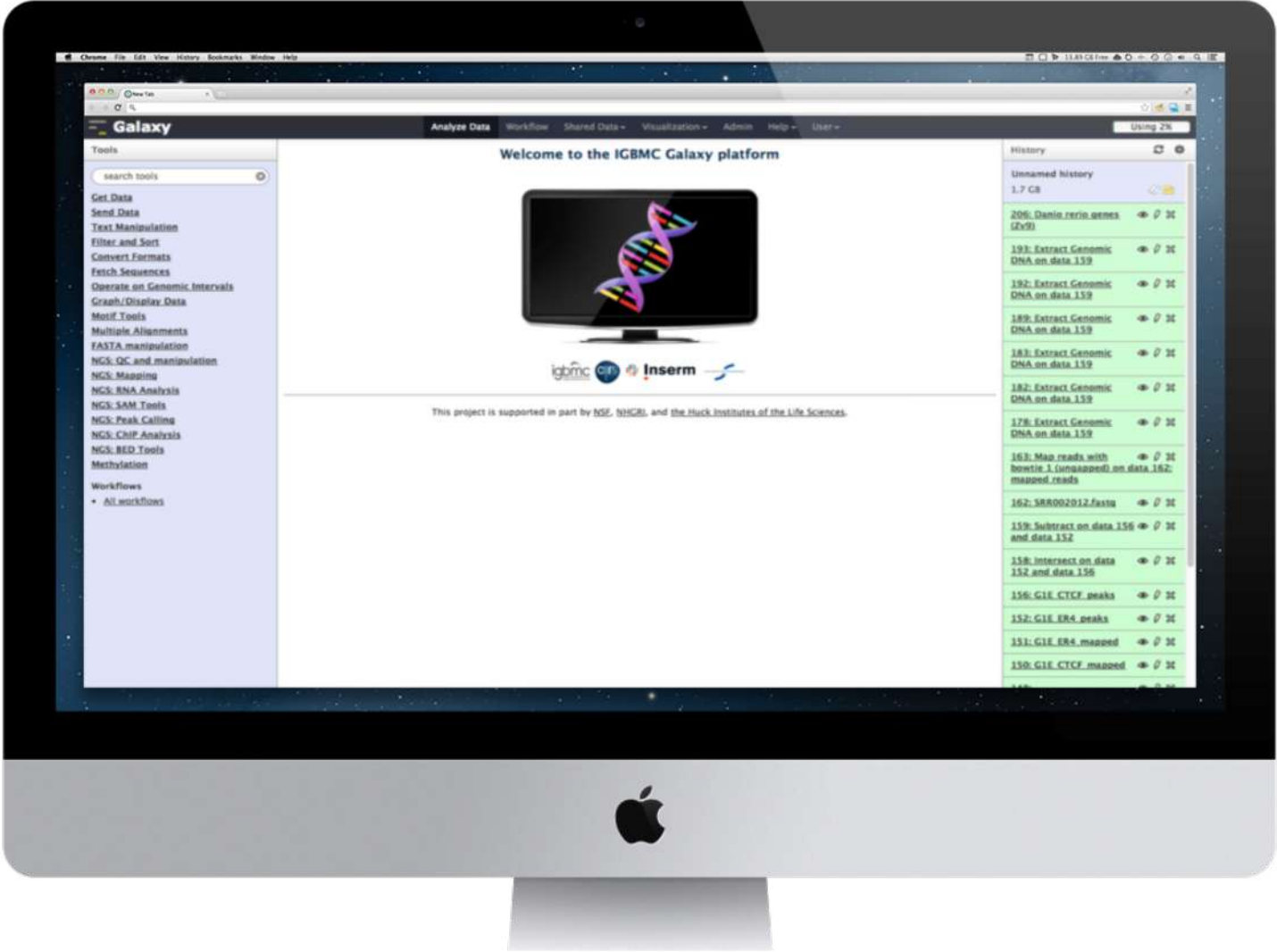
Galaxy is a **computing platform** that enables people to **run complex bioinformatics tools** on a **compute cluster** through a **simple web interface**.



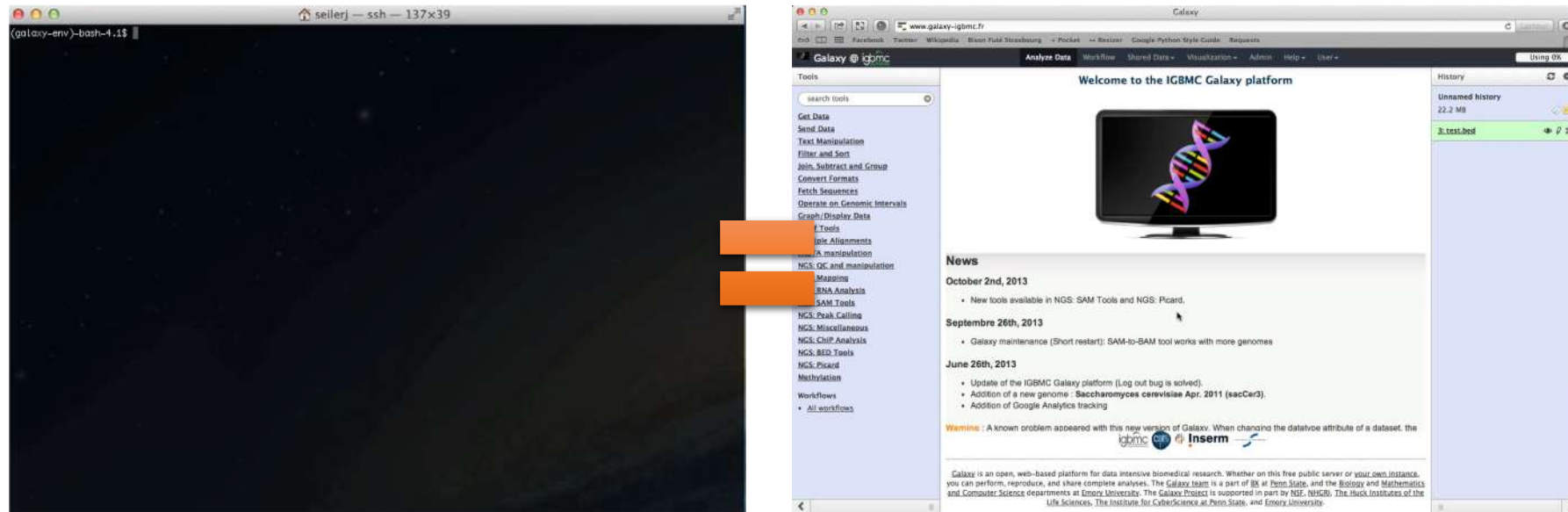
EMORY  
UNIVERSITY



# This is Galaxy



# Running analyses with tools



The image displays two side-by-side screenshots. The left screenshot shows a terminal window with the prompt `(galaxy-env)-bash-4.1$` and the user `seilerj` on host `ssh - 137x39`. The right screenshot shows a web browser displaying the Galaxy IGBMC platform. The browser address bar shows `www.galaxy-igbmc.fr`. The page title is "Welcome to the IGBMC Galaxy platform". The main content area features a large image of a DNA double helix on a monitor. Below the image is a "News" section with three entries: "October 2nd, 2013" (New tools available in NGS: SAM Tools and NGS: Picard), "September 26th, 2013" (Galaxy maintenance (Short restart): SAM-to-BAM tool works with more genomes), and "June 26th, 2013" (Update of the IGBMC Galaxy platform (Log out bug is solved), Addition of a new genome: *Saccharomyces cerevisiae* Apr. 2011 (sacCer3), Addition of Google Analytics tracking). A "Warnings" section below the news states: "A known problem appeared with this new version of Galaxy. When changing the database attribute of a dataset, the". The footer of the page contains the text: "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 [IB](#) at [Peru State](#), and the [Biology and Mathematics](#) and [Computer Science](#) departments at [Emory University](#). The Galaxy Project is supported in part by [NIH](#), [NH&MBS](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Peru State](#), and [Emory University](#)".

# Galaxy philosophy

- **Perform, and share** complete analysis
- **No programming skills** required
- **Open source** and **free** solution
- **Very large** and **active** community
- **Reproducibility/Usability/Transparency**

# How to use Galaxy

# Use Galaxy

- **Public servers**
- **Local servers**
- Clouds (Public, Commercial or Academic)
- Docker
- Virtual Machines



# Galaxy public servers

- Galaxy Project's public server (<https://usegalaxy.org/>) (3)
- There are several public remote Galaxy instances worldwide (160)
  - Genomics Servers
  - Domain Servers
  - Tool Publishing Servers

Public Galaxy Servers list :

<https://galaxyproject.org/use/>

Last Update on: 2020, December 16<sup>th</sup>

# Galaxy public servers

- All analyses are run on remote computing infrastructures
- No need to have a Supercomputer to use Galaxy
- Web browser



# Galaxy public servers

Your research institute



Connect to Galaxy web site through  
a web browser  
(<https://usegalaxy.org/>)

Download and  
upload of the data

Remote instance of Galaxy



Run analyses

# Galaxy public servers

Your research institute



World wide



Download and upload of the data



Remote instance of Galaxy



Uploading (takes a lot of time)  
Shared computing power  
(longer analysis)

# Galaxy local server

- Run a local production Galaxy because you want to
  - install and use tools unavailable on public Galaxies
  - use sensitive data (e.g. clinical)
  - process large datasets that are too big for public Galaxies
  - Develop Galaxy tools
  - Develop Galaxy itself



# Description of the main features of Galaxy

# Galaxy web interface

The image shows a screenshot of the Galaxy web interface. At the top, there is a dark blue navigation bar with the 'Galaxy' logo on the left and a menu on the right containing 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'Login or Register', and a grid icon. A 'Using 0%' indicator is on the far right. Below the navigation bar, the interface is divided into three main vertical sections. The left section is the 'Tools' panel, which has a search bar and lists various tool categories like 'Get Data', 'Collection Operations', 'GENERAL TEXT TOOLS', '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', and 'COMMON GENOMICS TOOLS'. The middle section is the main workspace, displaying a welcome message, a 'Galaxy Help' dialog box with the text 'Got Questions? Get Answers.' and the URL 'help.galaxyproject.org', and a 'Tweets' section featuring a tweet from '@galaxyproject' about group tags. The right section is the 'History' panel, which includes a search bar, the text 'Unnamed history (empty)', and a blue information box stating 'This history is empty. You can load your own data or get data from an external source'. Four orange callout boxes with arrows point to specific parts of the interface: 'Top menu' points to the navigation bar, 'Tool panel' points to the left sidebar, 'Data display and tools dialog window' points to the help dialog box, and 'History panel' points to the right sidebar.

# Top menu

The image shows a screenshot of the Galaxy web interface with several orange callout boxes and arrows pointing to specific menu items. The top navigation bar includes the following items: **Analyze Data**, **Workflow**, **Visualize**, **Shared Data**, **Help**, and **Login or Register**. The left sidebar contains a search bar and a list of tool categories: **Tools**, **Get Data**, **Collection Operations**, **GENERAL TEXT TOOLS**, **Text Manipulation**, **Filter and Sort**, **Join, Subtract and Group**, **Datamash**, **GENOMIC FILE MANIPULATION**, **FASTA/FASTQ**, **FASTQ Quality Control**, **SAM/BAM**, **BED**, and **VCF/BCF**. The main content area features a banner for "Try Galaxy on the Cloud" and a "Tweets by @galaxyproject" section. The right sidebar shows a "History" section and an "Unnamed" section.

Annotations with arrows pointing to the top menu:

- Run workflows** (points to the **Analyze Data** menu item)
- Run analyses** (points to the **Analyze Data** menu item)
- Access public data** (points to the **Shared Data** menu item)
- Log in/out, manage your account** (points to the **Login or Register** menu item)
- Get Help** (points to the **Help** menu item)



# Hands On

## Exercise 1

# History

The screenshot displays the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'Login or Register'. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: 'Get Data', 'Collection Operations', 'GENERAL TEXT TOOLS', and 'Text Manipulation'. The 'Text Manipulation' section lists several tools, including 'Compute an expression on every row'. The main workspace shows the configuration for the 'Compute an expression on every row' tool. The 'Add expression' field contains 'c3-c2'. The 'as a new column to' dropdown is set to 'No tabular dataset available'. The 'Round result?' dropdown is set to 'NO', and the 'Skip a header line' dropdown is set to 'no'. An 'Execute' button is visible. Below the tool configuration, there is a 'What it does' section and an 'Example' section. The right sidebar shows the 'History' panel, which is currently empty. A blue information box in the history panel states: 'This history is empty. You can load your own data or get data from an external source'. An orange box at the bottom right of the history panel contains the text: 'History panel' and 'Keep track of each job run'.

**Tools**

search tools

**Get Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

[Compute](#) an expression on every row

[annotateMyIDs](#) annotate a generic set of identifiers

[UniProt](#) ID mapping and retrieval

[Add input name as column](#) to an existing tabular file

[Replace column](#) by values which are defined in a convert file

[Replace](#) parts of text

[Text transformation](#) with sed

[Unfold](#) columns from a table

[Unique lines](#) assuming sorted input file

[Replace Text](#) in entire line

[Replace Text](#) in a specific column

[Multi-Join](#) (combine multiple files)

[Select last](#) lines from a dataset (tail)

[Cut](#) columns from a table (cut)

[Create text file with recurring lines](#)

**Compute an expression on every row (Galaxy Version 1.2.0)** Versions Options

**Add expression**

c3-c2

**as a new column to**

No tabular dataset available.

Dataset missing? See TIP below

**Round result?**

NO

**Skip a header line**

no

# characters are already considered as comments and kept

Execute

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

**What it does**

This tool computes an expression for every row of a dataset and appends the result as a new column (field).

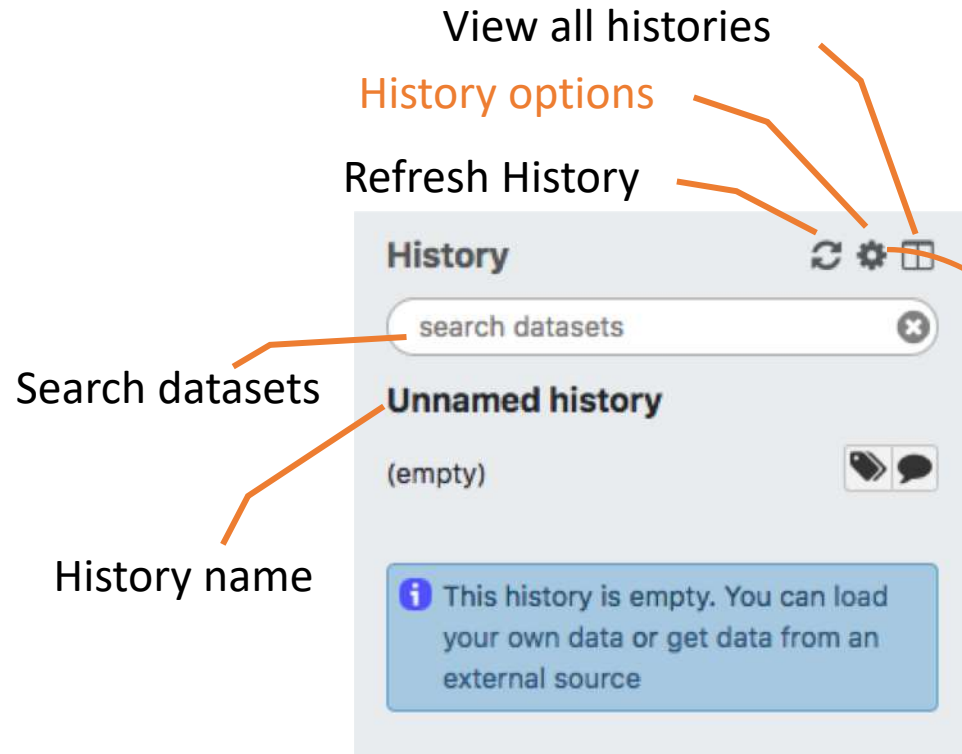
- Columns are referenced with **c** and a **number**. For example, **c1** refers to the first column of a tab-delimited file
- c3-c2** will add a length column to the dataset if **c2** and **c3** are start and end position

**Example**

If this is your input:

```
chr1 151077881 151077918 2 200 -
chr1 151081085 151082078 2 500 -
```

# History



## HISTORY LISTS

Saved Histories

Histories Shared with Me

## CURRENT HISTORY

Create New

Copy History

Share or Publish

Show Structure

Extract Workflow

Delete

Delete Permanently

Make Data Private

## DATASET ACTIONS

Copy Datasets

Dataset Security

Resume Paused Jobs

Collapse Expanded Datasets

Unhide Hidden Datasets

Delete Hidden Datasets

Purge Deleted Datasets

## DOWNLOADS

Export Tool Citations

Export History to File

## OTHER ACTIONS

Import from File

# View all histories

The screenshot displays the Galaxy web interface with a dark blue header. The header includes the Galaxy logo, navigation menus for 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User', and a 'Using 0%' indicator. Below the header, there are search bars for 'search histories' and 'search all datasets', along with a 'Create new' button. The main content area is divided into several vertical panels, each representing a 'Current History'. The first panel on the left is titled 'Unnamed history' and is empty, with a message 'This history is empty'. The second panel, also titled 'Unnamed history', shows 6 items with a total size of 127.21 KB. The items listed are: 1: lkpeaks\_notRBPJ +-150\_ran dom80\_adjacentSeq.fasta, 2: shuffleseq on data 1, 3: Galaxy14-[Intersect on data 13\_and data 1].bed, 4: Count on data 3, 5: TALs.fasta, and 6: L1spa\_ORF1.1.fastq. The third panel, titled 'Unnamed history', shows 2 items shown and 3 deleted, with a total size of 97.7 KB. The items are: 4: fimo.txt and 5: Correspondance\_JASPAR\_C ORE.txt. The fourth panel, titled 'Unnamed history', shows 1 item shown and 1 deleted, with a total size of 1.09 GB. The item is: 2: Brn2\_Day2\_rtta\_rep2.sort.bed. The fifth panel on the right is partially visible, titled 'phD', and shows 7 items with a total size of 126.01 GB. Each history panel includes a 'Switch to' button, a search bar for datasets, and icons for viewing, editing, and deleting items.

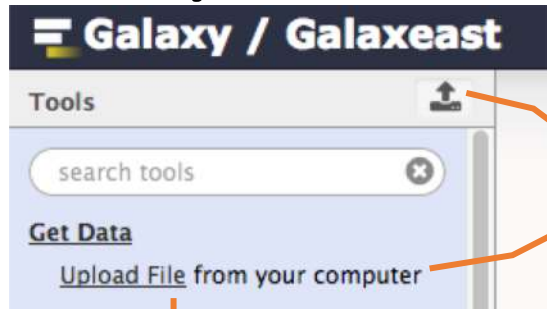
# Hands On

## Exercise 2

# Import data into Galaxy

- Your own data (from your computer)
- Shared data
- Data from external sources

# Import your own data to Galaxy





Display the drag and drop utility used to upload local files

## Download from web or upload from disk

Regular Composite

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
 Mouse_ChIP-Seq_example_Control_Data_chr19_mm9.fastq	84.1 MB	Auto-det...	unspecified (?)		

Name of the dataset

Size of the dataset

File format

Genome


Type (set all):

Auto-detect

Genome (set all):

unspecified (?)

 Choose local file

 Choose FTP file

 Paste/Fetch data

Pause

Reset

Start

Close

# Import shared data (data libraries)

Shared Data ▾ Visualiza (Top menu)

- Data Libraries
- Histories
- Workflows
- Visualizations
- Pages

Data Libraries

Galaxy / Galaxeast Analyze Data Workflow Shared Data ▾ Visualization ▾ Help ▾ User ▾ Using 0%

DATA LIBRARIES < 0 1 2 > showing 20 of 23 libraries Search Help

name↑	description	synopsis
<a href="#">Chip seq test dataset (chr10)</a>	MITF test dataset for chipSeq - chr10	MITF test dataset for chipSeq - chr10
<a href="#">Chip seq test dataset (chr2)</a>	Chip seq test dataset (chr2)	Chip seq test dataset (chr2)
<a href="#">Chromosome length</a>	*.len files for several assemblies	
<a href="#">Data demo AVIESAN</a>	Data demo AVIESAN	Data demo AVIESAN
<a href="#">Data Rustenholz</a>	Vitis vinifera data	Vitis vinifera data
<a href="#">Data_megadebug_jung</a>	Data_megadebug_jung	Data_megadebug_jung
<a href="#">Data_megadebug_velt</a>		
<a href="#">EBA2016</a>		
<a href="#">Fimo database (updated Mar. 16 2015)</a>		
<a href="#">Fimo databases</a>	Downloaded from MEME website	updated Jan 31 2013
<a href="#">Galaxy training - Chip seq test datasets</a>	Galaxy training - Chip seq test datasets	Galaxy training - Chip seq test datasets
<a href="#">Galaxy training - RNA seq test datasets</a>	Galaxy training - RNA seq test datasets	Galaxy training - RNA seq test datasets
<a href="#">Genome fasta files</a>	Genome fasta files (GATK - IGBMC)	Fasta files used by the IGBMC microarray and sequencing platform (to be used with GATK)
<a href="#">GTF</a>	Annotation files in GTF format	Annotation files in GTF format
<a href="#">Introduction 2 Galaxy (datasets)</a>	introduction 2 Galaxy (datasets)	
<a href="#">Jaspar motifs ID &lt;-&gt; name</a>	table of correspondences (for FIMO results...)	
<a href="#">MITF test dataset for chipSeq</a>	MITF test dataset for chipSeq	MITF (bam)
<a href="#">MITF test dataset for RNAseq</a>	MITF test dataset for RNAseq	
<a href="#">NGS course (Sep)</a>	NGS course (Sep)	NGS course (Sep)



# Import shared data (data libraries)

2. Import selected dataset to history

Galaxy / Galaxeast Analyze Data Workflow Shared Data Visualization Help User Using 0%

DATA LIBRARIES < 0 1 2 > showing 6 of 6 items  include deleted to History Download Delete Details Help

Libraries / Chromosome length

<input type="checkbox"/> name ↓	description	data type	size	time updated (UTC)	
<input type="checkbox"/> ..					
<input type="checkbox"/> <a href="#">ce10.len</a>		len	98 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">dm3.len</a>		len	227 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">hg19.len</a>		len	376 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">mm10.len</a>		len	1.4 KB	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">mm9.len</a>		len	330 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">tair10.len</a>		len	75 bytes	2015-01-08 01:25	

< 0 1 2 > showing 6 of 6 items

1. Select dataset

# Import public data

Tools

search tools

**Get Data**

- [Upload File](#) from your computer
- [UCSC Main](#) table browser
- [UCSC Archaea](#) table browser
- [EBI SRA](#) ENA SRA
- [BioMart](#) Ensembl server
- [GrameneMart](#) Central server
- [Flymine](#) server
- [modENCODE](#) fly server
- [modENCODE](#) modMine server
- [MouseMine](#) server
- [Ratmine](#) server
- [YeastMine](#) server
- [modENCODE](#) worm server
- [WormBase](#) server
- [ZebrafishMine](#) server
- [EuPathDB](#) server
- [GenomeSpace](#) import from file browser

Browse and import external data from public databases



Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

### Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the [OpenHelix Table Browser tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal genome: Mouse assembly: Dec. 2011 (GRCm38/mm10)

group: Genes and Gene Predictions track: UCSC Genes add custom tracks track hubs

table: knownGene describe table schema

region: genome position chr1:121427557-121432938 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

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

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

file type returned: plain text gzip compressed

get output summary/statistics

To reset all user cart settings (including custom tracks), [click here](#).

### Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

- clade:** Specifies which clade the organism is in.
- genome:** Specifies which organism data to use.

# Hands On

## Exercise 3.1

# Hands On

## Exercise 3.2

# Datasets/Jobs in the History

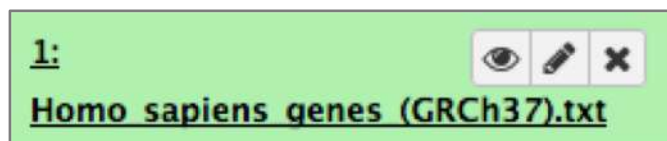
Grey: the job is waiting to run



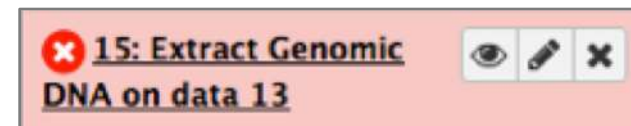
Yellow: the job is running



Green: the job is successfully done



Red: the job encountered a problem



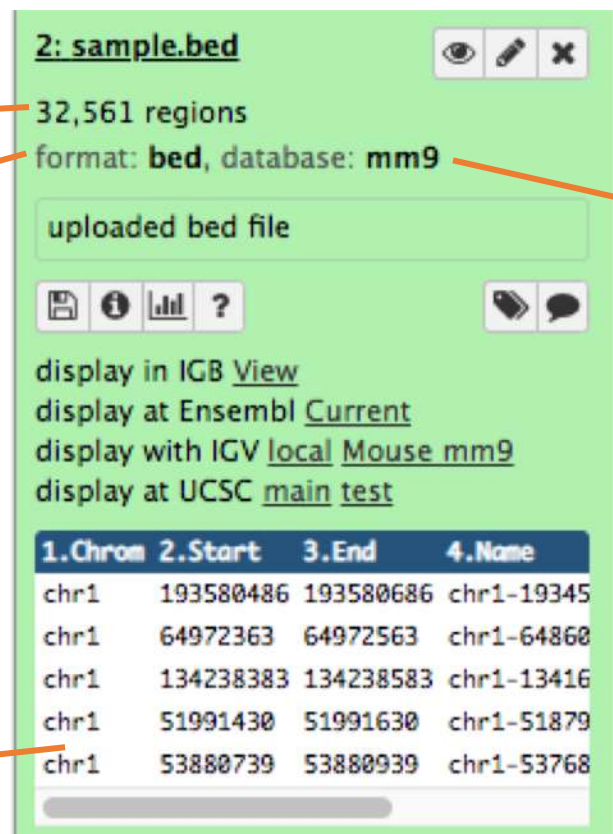
# Datasets/Jobs in the History




Number of lines  
in the file or size  
of the file

Format

Genome







If the dataset is a text  
file, the first lines of  
the file are displayed



**2: sample.bed**   

32,561 regions  
format: **bed**, database: **mm9**

uploaded bed file

display in IGB [View](#)  
display at Ensembl [Current](#)  
display with IGV [local](#) [Mouse](#) [mm9](#)  
display at UCSC [main](#) [test](#)

1.Chrom	2.Start	3.End	4.Name
chr1	193580486	193580686	chr1-19345
chr1	64972363	64972563	chr1-64860
chr1	134238383	134238583	chr1-13416
chr1	51991430	51991630	chr1-51879
chr1	53880739	53880939	chr1-53768

# Datasets/Jobs in the History

View dataset (if possible) in the middle panel of Galaxy

Download dataset

The screenshot shows a dataset viewer for '2: sample.bed'. At the top right, there are three icons: an eye (view), a pencil (edit), and an 'X' (delete). Below these, the text indicates '32,561 regions' and 'format: bed, database: mm9'. A text input field contains 'uploaded bed file'. Below this is a row of icons: a floppy disk (download), an information icon, a bar chart, and a question mark. To the right of these are a magnifying glass and a speech bubble. Below the icons, there are links for 'display in IGB View', 'display at Ensembl Current', 'display with IGV local Mouse mm9', and 'display at UCSC main test'. At the bottom, a table displays genomic data with columns for Chromosome, Start, End, and Name.

1. Chrom	2. Start	3. End	4. Name
chr1	193580486	193580686	chr1-19345
chr1	64972363	64972563	chr1-64860
chr1	134238383	134238583	chr1-13416
chr1	51991430	51991630	chr1-51879
chr1	53880739	53880939	chr1-53768

Delete dataset

Edit attributes of the dataset (change name, format, genome, permission)

# Size of histories and quota

The screenshot displays a software interface with a dark header bar at the top right showing a green progress indicator and the text "Using 20%". Below this is a "History" panel with a search bar labeled "search datasets". The panel lists "RNA-seq data analysis" with "2 shown" items. The first item is "7.23 GB" with a checkmark, a folder icon, and a speech bubble icon. Below it are two dataset entries: "2: sample.bed" and "1: siLuc3\_S12040.fastq", each with an eye icon, a pencil icon, and a close icon. An orange arrow labeled "Size of history" points to the "7.23 GB" text. Another orange arrow labeled "Quota" points to the "Using 20%" indicator.

Using 20%

History

search datasets

RNA-seq data analysis  
2 shown

7.23 GB

2: sample.bed

1: siLuc3\_S12040.fastq

Size of history

Quota



# Hands On

## Exercise 4

# Tool Panel / Run analyses

The screenshot displays the Galaxy web interface. On the left, the **Tools** panel is highlighted with an orange border and contains a search bar and a list of tool categories: **Get Data**, **Collection Operations**, **GENERAL TEXT TOOLS** (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), and **COMMON GENOMICS TOOLS**. An orange arrow points from the text "Tool panel" to this sidebar.

The central workspace features a navigation bar with "Analyze Data", "Workflow", "Visualize", "Shared Data", "Help", "Login or Register", and a "Using 0%" indicator. Below the navigation bar, a text block states: "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." A large black banner with white text reads "Galaxy Help Got Questions? Get Answers. help.galaxyproject.org". Below the banner is a "Tweets" section by @galaxyproject, featuring a tweet from Galaxy Project (@galaxyproject) about group tags for complex experimental designs, with a link to training.galaxyproject.org and the hashtag #usegalaxy.

On the right, the **History** panel shows a search bar for "search datasets" and indicates "Unnamed history (empty)". A blue information box states: "This history is empty. You can load your own data or get data from an external source".

# Tool Panel / Run analyses

Upload data

Search a tool

Tool category

Tools

search tools

Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Tools

search tools

Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Compute an expression on every row

annotateMyIDs annotate a generic set of identifiers

UniProt ID mapping and retrieval

Add input name as column to an existing tabular file

Replace column by values which are defined in a convert file

Tool

43

# Tools dialog window

**Galaxy** Analyze Data Workflow Visualize Shared Data Help Login or Register Using 0%

**Tools** search tools

**Get Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

- [Compute](#) an expression on every row
- [annotateMyIDs](#) annotate a generic set of identifiers
- [UniProt](#) ID mapping and retrieval
- [Add input name as column](#) to an existing tabular file
- [Replace column](#) by values which are defined in a convert file
- [Replace](#) parts of text
- [Text transformation](#) with sed
- [Unfold](#) columns from a table
- [Unique lines](#) assuming sorted input file
- [Replace Text](#) in entire line
- [Replace Text](#) in a specific column
- [Multi-Join](#) (combine multiple files)
- [Select last](#) lines from a dataset (tail)
- [Cut](#) columns from a table (cut)
- [Create text file with recurring lines](#)

**Compute an expression on every row (Galaxy Version 1.2.0)** Versions Options

**Add expression**

c3-c2

**as a new column to**

No tabular dataset available. Dataset missing? See TIP below

**Round result?**

NO

**Skip a header line**

no

# characters are already considered as comments and kept

Execute

**TIP:** If you have a dataset that is not tabular, you can use the [Convert](#) tool to convert it to a tabular file.

**What it does**

This tool computes an expression on every row of a dataset. The result is a new column in the dataset.

- Column name:
- Column name:

**Example**

If this is your input dataset:

```
chr1 1510 1510 1510 1510
```

## Data display and tools dialog window

- Set parameters
- Run tools
- Get help on tools
- Display content of dataset

# Hands On

## Exercise 5

# Hands On

## Exercise 6

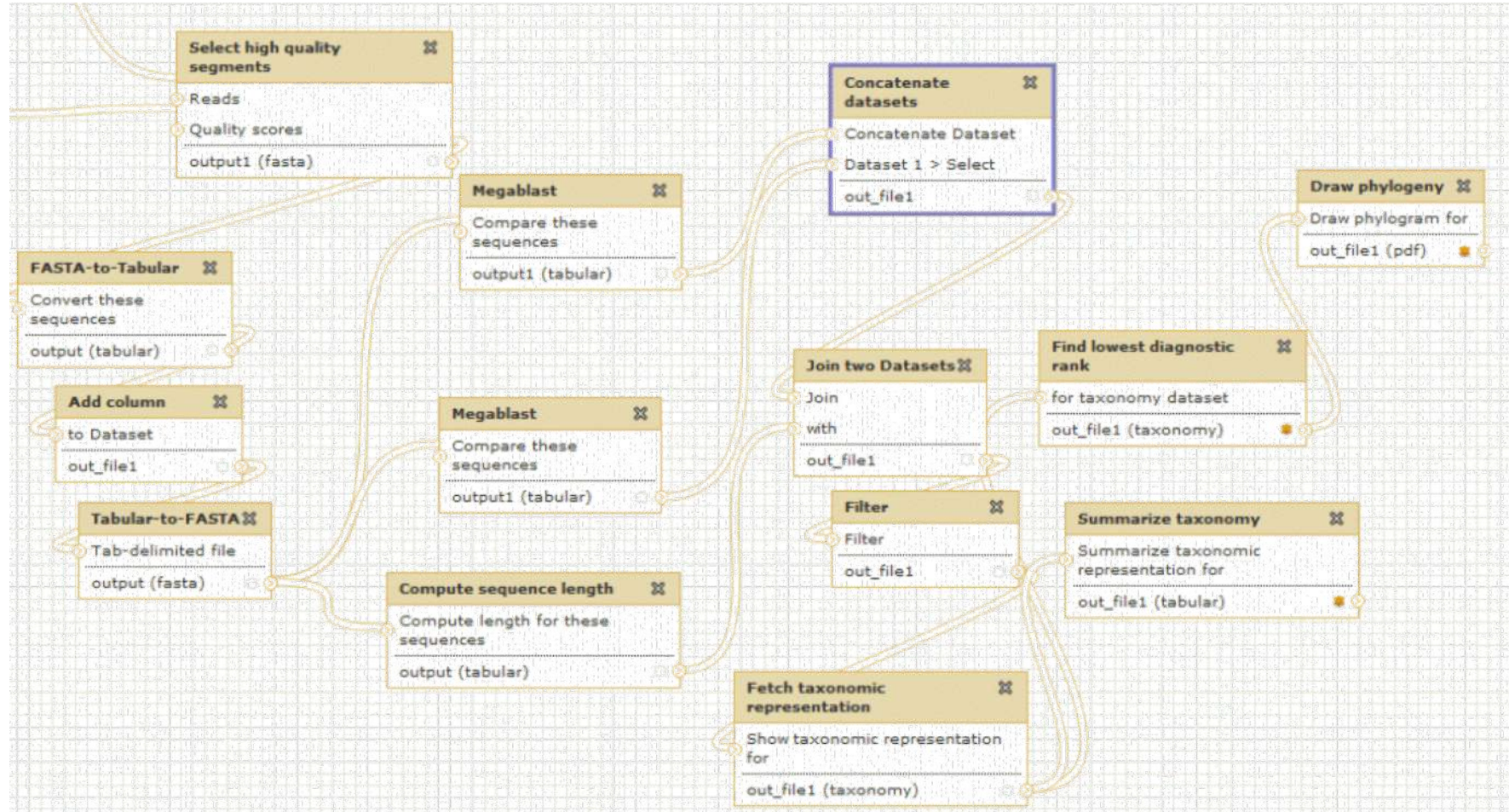
# Workflow

# What if we'd mix all together





# Galaxy workflow



# Galaxy workflows

- Workflow:
  - Analysis protocol with several steps (tools)
  - The output of a step is used as the input of the next next so file formats between two steps should be compatible!
- Workflows are often made general so that they can be run on various datasets
- Some of the parameters are pre-defined while others are set at runtime

# Workflows

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar lists various tool categories, with 'Workflows' and 'All workflows' at the bottom. The main content area features a 'Public Galaxy Servers and still counting' banner with a '080+' logo. Below the banner is a tweet from the Galaxy Project. The right sidebar shows the 'History' panel, which is currently empty.

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](#).

080+  
Public Galaxy Servers  
and still counting

Tweets by @galaxyproject

Galaxy Project @galaxyproject  
Did we mention: Galaxy Admin Training early registration ENDS IN 12 HOURS. [bit.ly/gat2016](http://bit.ly/gat2016)

History  
search datasets  
Unnamed history  
0 b  
This history is empty. You can [load your own data](#) or [get data from an external source](#)

Tools  
NGS: SAMtools  
NGS: BamTools  
NGS: Picard  
NGS: VCF Manipulation  
NGS: Peak Calling  
NGS: Variant Analysis  
NGS: RNA Structure  
NGS: Du Novo  
NGS: Gemini  
Operate on Genomic Intervals  
Statistics  
Graph/Display Data  
CloudMap  
Phenotype Association  
BEDTools  
Genome Diversity  
EMBOSS  
Regional Variation  
FASTA manipulation  
Multiple Alignments  
Metagenomic Analysis  
Multiple regression  
Multivariate Analysis  
Motif Tools  
STR-FM: Microsatellite Analysis  
NCBI SRA Tools  
DEPRECATED  
NGS: GATK Tools (beta)  
Workflows  
▪ All workflows

Create, run,  
edit (...)  
workflows

Run workflows

# Workflows

## Your workflows

You have no workflows.

## Workflows shared with you by others

No workflows have been shared with you.

## Other options

Configure your workflow menu

Create new workflow

Upload or import workflow

Create workflows

Create New Workflow

Workflow Name:

Workflow Annotation:

A description of the workflow; annotation is shown alongside shared or published workflows.

Create

Give a name to the workflow



# Workflow creation

Galaxy / Galaxeast Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools Workflow Canvas | Test Details

search tools

**Inputs**  
Get Data  
Send Data  
Text Manipulation  
Convert Formats  
Filter and Sort  
Join, Subtract and Group  
Extract Features  
Fetch Sequences  
Statistics  
Graph/Display Data

NGS TOOLBOX BETA  
NGS: QC and manipulation  
NGS: SAM Tools  
Operate on genomic intervals  
Motif tools  
FASTA manipulation  
NGS: GATK Tools (beta)  
NGS: Peak Calling  
NGS: Homer  
NGS: BEDtools  
NGS: Picard  
NGS: Variant Annotation  
NGS: Miscellaneous  
NGS: RNA Analysis  
NGS: Mapping  
NGS: DeepTools  
NGS: RSeQC  
Multiple alignments

**Edit Workflow Attributes**

**Name:**  
Test

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

**Annotation / Notes:**  
test  
Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.

Add tools or input datasets to the workflow

# Workflow creation

The screenshot displays the Galaxy workflow editor interface. On the left, a 'Tools' sidebar lists various categories such as 'Inputs', 'Get Data', 'Send Data', 'Text Manipulation', 'Convert Formats', 'Filter and Sort', 'Join, Subtract and Group', 'Extract Features', 'Fetch Sequences', 'Statistics', 'Graph/Display Data', and 'NGS TOOLBOX BETA'. The 'Filter and Sort' category is expanded, showing several options for filtering data. The central 'Workflow Canvas' shows a workflow with two tools: 'Input dataset' (output: 'output') and 'Filter' (output: 'out\_file1'). The 'Filter' tool is selected, and its configuration details are shown in the 'Details' panel on the right. The details panel includes a dropdown menu for 'Filter data on any column using simple expressions (Galaxy Version 1.1.0)', a 'Filter' section with a text input field containing 'c1==chr22', a 'Number of header lines to skip' field set to '0', and sections for 'Email notification' and 'Output cleanup'.

Input dataset.

Most of the time, a workflow starts with an input dataset to which analyses are applied. In Galaxy, the file format of the input dataset will be limited to the input file format of the subsequent step

Tool to be run

# Workflow creation

The screenshot displays the Galaxy/Galaxeast workflow creation interface. The main canvas shows two steps: 'Input dataset' and 'Filter'. A green line connects the 'output' of the 'Input dataset' step to the 'Filter' step. An orange arrow points from the text below to this green link. The 'Filter' step is configured with the condition 'c1=='chr22'' and 'Number of header lines to skip' set to 0. The 'Details' panel on the right shows the configuration for the 'Filter' tool.

If two steps can be linked together, the link between the two boxes is green

# Workflow creation

The screenshot shows the Galaxy / Galaxeast interface for creating a workflow. The 'Tools' panel on the left lists various tool categories, including 'Filter and Sort'. The 'Workflow Canvas' in the center displays a workflow with two steps: 'Input dataset' and 'Filter'. The 'Filter' step is selected, and its configuration panel is open on the right. The configuration panel includes a dropdown menu for 'Filter data on any column', a text input field for 'With following condition' containing 'c1=='chr22'', and a text input field for 'Number of header lines to skip' set to '0'. There are also sections for 'Annotation / Notes', 'Email notification', and 'Output cleanup'.

Pre-configure tool parameters and configure parameters to be set at run time



# Workflow creation

The screenshot shows the Galaxy workflow editor interface. On the left is a 'Tools' sidebar with a search bar and various tool categories like 'Inputs', 'Text Manipulation', and 'NGS TOOLBOX BETA'. The central 'Workflow Canvas' shows a workflow with a 'Filter' tool connected to a 'Sort' tool. A tooltip is visible over the 'Filter' tool, stating: 'Mark dataset as a workflow output. All unmarked datasets will be hidden.' An orange arrow points from a star icon on the 'Filter' tool to this tooltip. Another orange arrow points from a star icon on the 'Sort' tool to a 'Configure Output' dialog box for 'out\_file1'.

Click on star to select which datasets will be displayed in the history generated when running of the workflow

Click to get the parameter to be set at runtime

# Workflow creation

Save, run workflows

The screenshot displays the Galaxy / Galaxeast workflow editor. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The main area is the 'Workflow Canvas | Test', which shows a workflow with two tools: 'Filter' and 'Sort Dataset'. The 'Filter' tool is connected to the 'Sort Dataset' tool. A context menu is open over the 'Filter' tool, listing options: 'Save', 'Run', 'Edit Attributes', 'Auto Re-layout', and 'Close'. An orange arrow points from the text 'Save, run workflows' to the 'Run' option in the menu. On the left, a 'Tools' sidebar lists various categories like 'Inputs', 'Get Data', 'Send Data', etc. On the right, a 'Details' panel is visible, showing configuration options for the selected tool, including a condition 'c1=='chr22'' and a 'Number of header lines to skip' set to 0.

# Run workflows

Set input file(s)

The screenshot displays the Galaxy/Galaxeast interface for running a workflow named "chip workflow". The interface is organized into several panels:

- Tools Panel (Left):** A sidebar containing various tool categories such as "Get Data", "Send Data", "Text Manipulation", "Convert Formats", "Filter and Sort", "Join, Subtract and Group", "Extract Features", "Fetch Sequences", "Statistics", "Graph/Display Data", and "NGS TOOLBOX BETA".
- Workflow Configuration Panel (Center):** Titled "Running workflow 'chip workflow'", it shows four steps:
  - Step 1: Input dataset:** The "Input Dataset" field is set to "4: chr10\_ctr2\_1.fastq.gz".
  - Step 2: Map with Bowtie for Illumina (version 1.1.3):** This step is currently collapsed.
  - Step 3: MACS (version 1.4.2):** This step is currently collapsed.
  - Step 4: homer annotatePeaks (version 0.0.5):** The "Homer peaks OR BED format" is set to "Output dataset 'output\_bed\_file' from step 3". The "Genome version" dropdown is set to "tair10". The "Extra options" checkbox is checked. The "Action" is set to "Hide output 'out\_log'".
- History Panel (Right):** Titled "History", it shows a search bar and a list of datasets. The current dataset is "4: chr10\_ctr2\_1.fastq" with a size of 120.7 MB and format "fastqsanger, database: hg19".

Three orange arrows point to specific elements in the interface:

- An arrow points from the text "Set input file(s)" to the "Input Dataset" field in Step 1.
- An arrow points from the text "Set parameters" to the "Genome version" dropdown in Step 4.
- An arrow points from the text "Run workflow" to the "Run workflow" button at the bottom of the configuration panel.

# Hands On

## Exercise 7

# Hands On

## Exercise 8

# Hands On

## Exercise 9

# Privacy

- By default datasets, workflows, histories are private to the user that generated/uploaded them.
- They can be shared across Galaxy users (of the same Galaxy instance) or via links