

# Data Platforms for Life Science: Galaxy

---

2018 SACNAS

October 10, 2018

San Antonio, Texas, United States

Dave Clements

Johns Hopkins University

Galaxy Team / Galaxy Community



**#usegalaxy @galaxyproject**

# Links

- These slides
  - [bit.ly/gxy-sacnas-2018](https://bit.ly/gxy-sacnas-2018)
- Peaks to Genes Tutorial
  - [bit.ly/gxy-p2g](https://bit.ly/gxy-p2g)
- European UseGalaxy Server
  - [usegalaxy.eu](https://usegalaxy.eu)
- Network
  - [SACNAS2018 / 2018sacnas](#)

# What is Galaxy?

Keith Bradnam's definition:

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

From

"13 Questions You May Have About Galaxy"

<http://bit.ly/13questions>

**Galaxy is available several ways ...**

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



# As a free for everyone service on the web:

## usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

Galaxy

Analyze DataWorkflowVisualizeShared DataHelpLogin or RegisterUsing 0%

Tools

search tools

Get Data

Lift-Over

Collection Operations

Text Manipulation

Datamash

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

NGS: QC and manipulation

NGS: DeepTools

NGS: Mapping

NGS: RNA Analysis

NGS: SAMtools

NGS: BamTools

NGS: Picard

NGS: VCF Manipulation

NGS: Peak Calling

NGS: Variant Analysis

NGS: RNA Structure

NGS: Du Novo

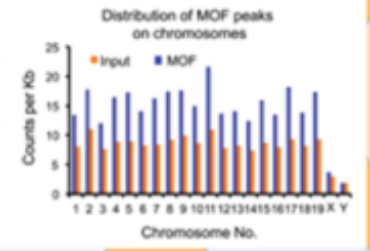
NGS: Gemini

NGS: Assembly

NGS: Chromosome Conformation

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](#).

From Peaks to Genes  
an introduction to Galaxy tutorial



Galaxy Training NetworkLi et al., Cell Stem Cell, 2012


Tweets by @galaxyproject

Galaxy Project Retweeted

The Carpentries

@thecarpentries

Attending #SACNAS2018 in San Antonio, TX & looking to get more from your data? Still spots in the Oct 10 pre-conference @swcarpentry @datacarpentry workshop & @cyverse @galaxy Data Platforms for Life Sciences workshop. Register here! [sacnas-data.eventbrite.com](#)



PennState

JOHNS HOPKINS UNIVERSITY

OREGON HEALTH & SCIENCE UNIVERSITY

Galaxy / Australia

Analyze DataWorkflowVisualizeShared DataHelpLogin or RegisterUsing 0%

Tools

search tools

FILE AND META TOOLS

Get Data

Convert Formats

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Extract Features

Fetch Sequences

Fetch Alignments

Galaxy Australia

Welcome to Galaxy Australia

Galaxy Australia is currently running Galaxy version 18.05

Galaxy is a web-based platform for data intensive biological research.

Users without programming experience can specify parameters and run tools and workflows. Galaxy also automatically captures information so that any user can repeat and understand a complete computational analysis.

Service is free to use for any Australian researcher. [On-line](#) training material is available to help get you started and for the most popular analysis methods.

[Data Storage Policy](#) \*\*JULY 2018 news – note new data limits and retention times

[Get help for installation of software tool or reference dataset on Galaxy Australia](#)

This public Galaxy Service is provided to you by:

Galaxy / Europe

Analyze DataWorkflowVisualizeShared DataHelpLogin or RegisterUsing 0 bytes

Tools

search tools

FILE AND META TOOLS

Get Data

Convert Formats

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMICS, NGS

Extract Features

BED Tools

Fetch Alignments

Operate on Genomic Intervals

Multiple Alignments

FASTA/FASTQ manipulation

Picard

Quality Control

Assembly

Mapping

Variant Calling

Genome editing

GATK Tools

Gemini Tools

RNA Analysis

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

News

Oct 8, 2018

Initial release of `gxadmin` tool

Oct 2, 2018

Tutorial of the Month: Maria Doyle selected "From peaks to genes"

Sep 24, 2018

A successful Galaxy HTS data analysis workshop

Sep 21, 2018

UseGalaxy.eu Downtime

Jul 25, 2018

Freiburg Galaxy Team has an open position!

Jul 18, 2018

GCCBOSC

Events

Dec 6, 2018 - Dec 7, 2018

Galaxy RNA-seq data analysis workshop in Freiburg

Nov 12, 2018 - Nov 16, 2018

Elixir Biohackathon 2018

Oct 17, 2018 - Oct 19, 2018

Bioinformatics for Translational Medicine using Galaxy: see it, do it, teach it!

Sep 17, 2018 - Sep 21, 2018

Galaxy HTS data analysis workshop in Freiburg

Sep 4, 2018

RNA-Seq data analysis with Galaxy for clinical applications

Aug 27, 2018 - Aug 29, 2018

Next Generation Sequencing data analysis, LUMC

Training Infrastructure as a Service

We want to help you conduct your training seminars. You provide the training.

OPEN CHAT

# UseGalaxy.eu

Let's take a look

# Learning Galaxy

Lots of ways to learn. Here are two:

- Workshops / training (we are here!)  
[galaxyproject.org/events/](https://galaxyproject.org/events/)
- Galaxy Training Network Tutorials (we will do one!)  
[training.galaxyproject.org/](https://training.galaxyproject.org/)

# Intro to Galaxy: Peaks to Genes Tutorial

**[bit.ly/gxy-p2g](https://bit.ly/gxy-p2g)**



# Ways to Use Galaxy: Academic Clouds

Launch your own Galaxy server on a compute cloud

- Academic clouds are typically free to use, after an application
- Usually limited by geography
- Easily launch a Galaxy instance, just for you or your group, and run it only as long as you need it

Galaxy is available on academic clouds in the **US**, Canada, Australia, UK, Poland, Belgium, Sweden, Italy ...



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

# Academic Clouds: Jetstream (US)

- Requires an XSEDE allocation
- Faculty can apply
  - Can then share allocation with their colleagues and students
- XSEDE *really, really* wants to help you do your research



XSEDE

Extreme Science and Engineering  
Discovery Environment

[jetstream-cloud.org](http://jetstream-cloud.org)   [xsede.org](http://xsede.org)

# Ways to Use Galaxy: Public Galaxy servers

## General Purpose / Genomics

- Wide range of genomic analysis tools
- 18 servers (incl. 3 UseGalaxy)

## Domain Specific

- Proteomics, viruses, metagenomics, ChIP-Seq, plants, image analysis, natural language processing, ...
- 39 servers

## Tool publishing servers

- Groups use Galaxy to make it easy to use their tools on the web
- 43 servers

## Total

100 servers

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

# Ways to Use Galaxy: Commercial Cloud Providers



<http://aws.amazon.com/education>

<http://globus.org/>

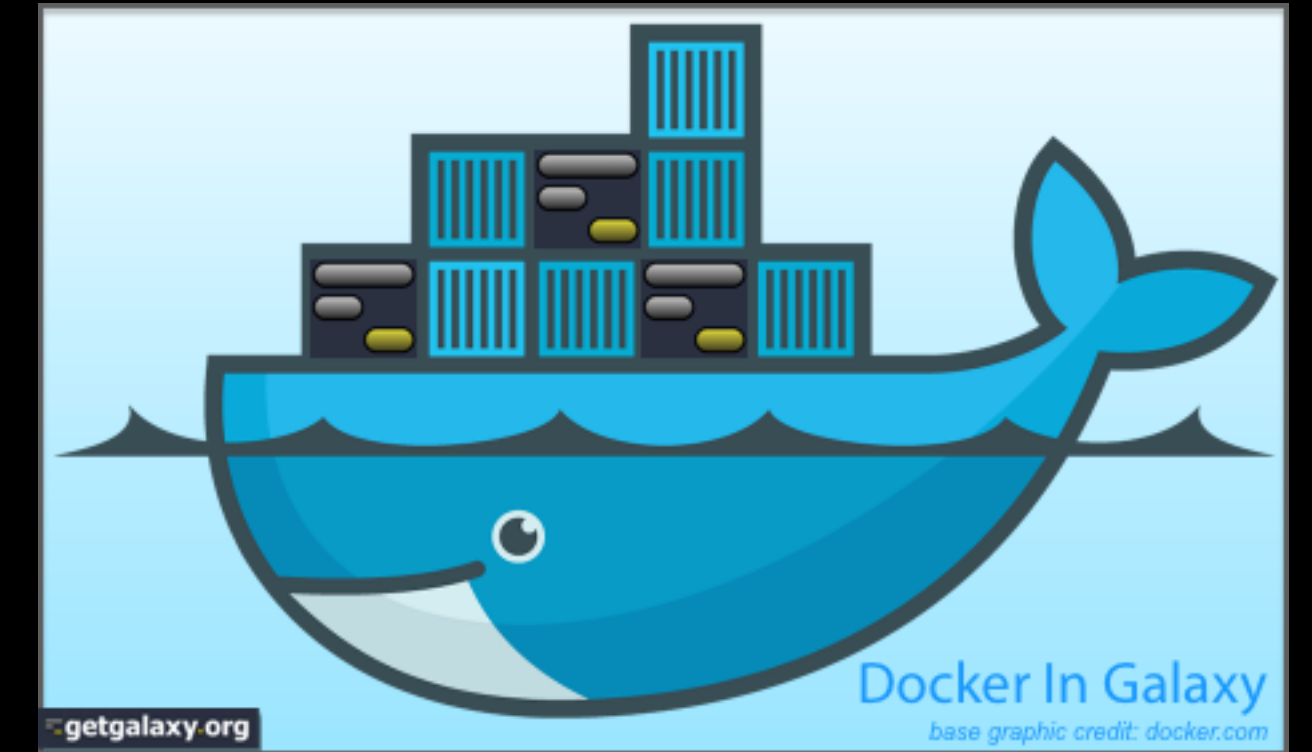
<https://launch.usegalaxy.org/>

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



# Ways to Use Galaxy: VMs & Containers

- Preconfigured Galaxy Virtual Machines (VMs) and Docker Containers can be downloaded and run on your laptop or lab server
- Minimal setup and configuration.
- Uses your hardware.



### Containers

These Galaxy instances are prepackaged using container technology, usually Docker. You run these locally by first installing the container technology (Docker is easy), and then launching the containerized Galaxy within that technology. These use local resources.

Show  entries

Search:

Resource	Summary	Keywords
<a href="#">ballaxy</a>	Hosts the <a href="#">BALL (Biochemical Algorithms Library) Project tools</a> , i.e. computer aided drug design and molecular modelling based on protein and ligand structure data.	<a href="#">Domain</a>
<a href="#">CRAVAT-P</a>	A Docker image containing a fully-operational Galaxy instance with pre-installed demonstration material for CRAVAT-P.	<a href="#">Tools</a>
<a href="#">IRProfiler</a>	Immune Repertoire Profiler (IRProfiler), a novel software pipeline that delivers a number of core receptor repertoire quantification and comparison functionalities on high-throughput TR and BcR sequencing data.	<a href="#">Tools</a>
<a href="#">RiboGalaxy</a>	Tools for the analysis and visualization of ribo-seq data obtained with the ribosome profiling technique.	<a href="#">Domain</a>
<a href="#">RNA Workbench</a>	more than 50 bioinformatics tools dedicated to different research areas of RNA biology, including RNA structure analysis, RNA alignment, RNA annotation, RNA-protein interaction, ribosome profiling, RNA-Seq analysis, and RNA target prediction.	<a href="#">Domain</a>
<a href="#">RNAcommender</a>	RNAcommender is a tool for genome-wide recommendation of RNA-protein interactions. It is a recommender system capable of suggesting RNA targets to unexplored RNA binding proteins, by propagating the available interaction information, taking into account the protein domain composition and the RNA predicted secondary structure.	<a href="#">Tools</a>
<a href="#">SMAGEXP</a>	SMAGEXP (Statistical Meta-Analysis for Gene EXPression) provides a unified way to carry out meta-analysis of gene expression data, while taking care of their specificities.	<a href="#">Tools</a>

### Virtual Machines (VMs)

VMs are similar to containers but use a different technology. Containers take significant advantage of your computer's underlying operating system. Virtual Machines include an entire supporting operating system, and are significantly larger than containers. You run these locally by first installing a VM player like VirtualBox, and then downloading and launching the VM within that player. These use local resources.

Show  entries

Search:

Resource	Summary	Keywords
<a href="#">DDBJ Read Annotation Pipeline</a>	Annotates raw sequencing reads from next-generation sequencers(NGS) with high throughput, which are registered in DDBJ Read Archive (DRA).	<a href="#">Tools</a>
<a href="#">ImmunoGlobulin Galaxy (IGGalaxy)</a>	Provides a set of tools for easy detection and quantitation of immunoglobulin heavy chain alternative transcripts.	<a href="#">Domain</a>
<a href="#">KREAP</a>	Kinetic Re-Epithelialization Analysis Pipeline	<a href="#">Tools</a>
<a href="#">Pitagora-Galaxy</a>	The <a href="#">Pitagora-Galaxy Project</a> server intended for testing and sharing. Heavy analysis should be performed using the project's identical virtual machine (VM) or Amazon Machine Image (AMI).	<a href="#">Genomics</a>

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

# Ways to use Galaxy: local installs

- Galaxy is open-source. Anyone can install it at their institution.
- Works on platforms from laptops to supercomputers.

[getgalaxy.org](https://getgalaxy.org)





# Galaxy Community and Ecosystem

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

# Support? Questions?

[galaxyproject.org/search](https://galaxyproject.org/search)  
[gitter.im/galaxyproject/Lobby](https://gitter.im/galaxyproject/Lobby)  
[biostar.usegalaxy.org](https://biostar.usegalaxy.org)

[www.biostars.org/](https://www.biostars.org/)  
(general bioinformatics)

# Acknowledgements

You

Gina Wall

Blake Joyce Tracy Teal Camille Avestruz Joslynn Lee  
Bérénice Batut Saskia Hiltemann Helena Rasche Björn Grüning  
SACNAS NIH AnVIL



Galaxy Team and Community

Johns Hopkins University

Penn State University

Oregon Health & Science University

Lerner Research Institute Cleveland Clinic

A large, colorful mural of a thumbs-up emoji is painted on a wall. The emoji is yellow with a wide smile and a white hand giving a thumbs-up. The background of the mural is green and blue. The word "Thanks" is written in white text across the top of the image.

# Thanks

Dave Clements

Galaxy Project

Johns Hopkins University

[clements@galaxyproject.org](mailto:clements@galaxyproject.org)