

# Introduction to Galaxy

---

## Cal State San Marcos

January 12, 2018

San Marcos, California, United States

Dave Clements

Johns Hopkins University

Galaxy Team / Galaxy Community



#usegalaxy @galaxyproject

# Agenda

9:00 Welcome: Introductions, logistics, background

## **Galaxy with RNA-seq**

9:30 Concepts, formats, and quality control

10:45 Break

11:00 Mapping reads to a genome

12:20 Lunch

1:20 Differential expression

2:50 Break

3:05 Creating reusable workflows

5:00 Done

# Today's materials

- Some slides specific to this workshop
- Galaxy Training Network
  - Slides and Tutorial
- All of it available online.
  - <https://galaxyproject.org/events/2018-01-csusm-workshop/>
  - <http://bit.ly/gxy-sanmarcos>

# Today's Goals

- Teaching you how to do bioinformatic analysis in Galaxy
- We'll use an RNA-Seq example to do that
- At the end of the day you'll have a basic understanding of
  - how to use Galaxy to explore, run, and share bioinformatics analyses
  - how to get Galaxy and bioinformatics help
  - RNA-Seq data analysis

# 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 User Interface

- Lets do a whirlwind introduction to the Galaxy interface.
- Most of what we touch in this intro we will revisit later today.

[usegalaxy.org](http://usegalaxy.org)

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

<http://galaxyproject.org>

# As a free for everyone service on the web: [usegalaxy.org](http://usegalaxy.org)

The screenshot displays the Galaxy web interface. On the left is a navigation menu with categories such as 'Get Data', 'Text Manipulation', 'NGS: QC and manipulation', and 'NGS: RNA Analysis'. The main content area features a header with navigation tabs like 'Analyze Data', 'Workflow', and 'Shared Data'. Below this is a text block describing Galaxy as an open-source platform for biomedical research, with links to 'start here', 'help resources', 'tutorial', and 'Tool Shed'. A central graphic for 'GAME 2017 Melbourne 3-9 February' includes a red globe icon and the text 'Talk abstracts due 30 November'. To the right, a Twitter feed shows tweets from the Galaxy Project and Pratik Jagtap. The footer contains logos for Penn State, Johns Hopkins, TACC, and Cyverse.


**Galaxy** Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools

search tools

[Get Data](#)  
[Lift-Over](#)  
[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](#)  
[Operate on Genomic Intervals](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Phenotype Association](#)


**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](#).

 **GAME**  
**2017**  
Melbourne  
3-9 February

**Talk abstracts due 30 November**

Tweets by @galaxyproject

Galaxy Project Retweeted

 **Pratik Jagtap** @pratkomics  
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

PENNSSTATE  
JOHNS HOPKINS  
TACC  
CYVERSE



# Public Galaxy servers

## General Purpose / Genomics servers

19 servers as of this yesterday

## Domain Specific

36 servers

## Tool publishing servers

44 servers

## Total

99 servers

[bit.ly/gxyServers](http://bit.ly/gxyServers)

# Galaxy Services



# Can't use a *service*? Galaxy is available on other Clouds



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

<http://globus.org/>

<http://wiki.galaxyproject.org/Cloud>

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



# Hands On: RNA-Seq in Galaxy

<http://galaxyproject.github.io/training-material/topics/transcriptomics/tutorials/srna/tutorial.html>

<http://bit.ly/csusm-rna>

[bit.ly/gxy-sanmarcos](http://bit.ly/gxy-sanmarcos)

# Community and Ecosystem

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

# The Galaxy Team



Enis Afgan

Dannon Baker

Dan Blankenberg

Dave Bouvier

Marten Cech

John Chilton



Dave Clements

Nate Coraor

Jeremy Goecks

Sam Guerler

Jen Jackson



Vahid Jalili

Delphine Lariviere

Anton Nekrutenko

Nick Stoler

James Taylor

<https://galaxyproject.org/galaxy-team>



# Acknowledgements

You

Arun Sethuraman

College of Science and Mathematics

Department of Biological Sciences

Cal State San Marcos

B er enice Batut   Saskia Hiltemann   Bj orn Gr uning

XSEDE & Jetstream

NIH   NSF

Johns Hopkins University

Penn State University

Oregon Health & Science University





**Dave Clements**

Galaxy Project

Johns Hopkins University

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

**Thanks**