

Utilizing the Galaxy Analysis Framework at Core Facilities

Western Association of Core Directors (WACD)
September 18, 2015

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



#usegalaxy @galaxyproject

Talk plan when I got off the train yesterday morning

- 1/3 What is Galaxy and what can it do?
- 1/3 Help clients to do their own data analysis w/ Galaxy!
- 1/6 Using Galaxy for in-house data pipelines
- 1/6 Q & A

<http://galaxyproject.org>

Talk plan after yesterday

- ~~1/3 What is Galaxy and what can it do?~~
- ~~1/3 Help clients to do their own data analysis w/ Galaxy!~~
- ~~1/6 Using Galaxy for in-house data pipelines~~
- ~~1/6 Q & A~~

- 2/10 What is Galaxy and what can it do?
- 1/10 Using Galaxy for in-house data pipelines
- 2/10 Help clients to do their own data analysis w/ Galaxy!
- 5/10 Open discussion:
 - What is the role of cores in supporting client data analysis?*
 - Should this be part of your value proposition?*

<http://galaxyproject.org>

What is Galaxy?

Data integration and analysis platform that emphasizes accessibility, reproducibility, and transparency

<http://galaxyproject.org>

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>

What: A web based platform

The screenshot displays the Galaxy web-based platform interface. The top navigation bar includes the Galaxy logo and menu items: Analyze Data, Workflow, Shared Data, Visualization, Cloud, Help, and User. A status bar on the right indicates 'Using 3%'. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: Get Data, Send Data, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, NGS: QC and manipulation, NGS: Mapping, NGS: RNA-seq, NGS: SAMtools, NGS: BAM Tools, NGS: Picard, NGS: VCF Manipulation, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, Operate on Genomic Intervals, Statistics, Graph/Display Data, Phenotype Association, snpEff, and BEDTools.

The main panel shows the configuration for the 'MACS Model-based Analysis of ChIP-Seq (Galaxy Tool Version 1.0.1)' tool. The configuration includes the following fields:

- Experiment Name:** MACS in Galaxy
- Paired End Sequencing:** Single End
- ChIP-Seq Tag File:** 6: Tags Chr19 SAM
- ChIP-Seq Control File:** 5: Control Chr19 SAM
- Effective genome size:** 1870000000 (default: 2.7e+9)
- Tag size:** 36
- Band width:** 300
- Pvalue cutoff for peak detection:** 0.00001 (default: 1e-5)
- Select the regions with MFOLD high-confidence enrichment ratio against background to build model:** (checkbox)

The right sidebar shows the 'History' panel with a search bar and a list of datasets. The datasets are listed in descending order of size, with the top entry being 'Basic Protocol 3: ChIP-Seq' (15 shown, 1.6 GB). The datasets are:

- 15: Filter pileup on data 14
- 14: Generate pileup on data 13: converted pile up
- 13: SAM-to-BAM on data 6: converted BAM
- 12: MACS on data 5 and data 6 (html report)
- 11: MACS on data 5 and data 6 (control: wig)
- 10: MACS on data 5 and data 6 (treatment: wig)

The bottom of the history panel shows a summary for the selected dataset (10: MACS on data 5 and data 6 (treatment: wig)): ~31,000,000 lines, format: wig, database: mm9. It also includes links to 'display in IGB View' and 'display at UCSC main'.

What: Data integration

Tools



Get Data

[Upload File](#) from your computer

[UCSC Main](#) table browser

[UCSC Archaea](#) table browser

[EBI SRA](#) ENA SRA

[BioMart](#) Central 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

Send Data

Workflow Shared Data Visualization Cloud Help User



Using 3%

based Analysis of ChIP-Seq (Galaxy Tool Version 1.0.1)

Options

ne

encing

File

6: Tags Chr19 SAM

ontrol File

5: Control Chr19 SAM

ne size

or peak detection

ns with MFOLD high-confidence enrichment ratio against background

History



Basic Protocol 3: ChIP-Seq

15 shown

1.6 GB



[15: Filter pileup on data 14](#)



[14: Generate pileup on data 13: converted pileup](#)



[13: SAM-to-BAM on data 6: converted BAM](#)



[12: MACS on data 5 and data 6 \(html report\)](#)



[11: MACS on data 5 and data 6 \(control: wig\)](#)



[10: MACS on data 5 and data 6 \(treatment: wig\)](#)



~31,000,000 lines

format: wig, database: mm9



display in IGB [View](#)

display at UCSC [main](#)

1

NGS: QC and manipulation

[FastQC Read Quality reports](#)

[Select high quality segments](#)

[Build base quality distribution](#)

[Draw quality score boxplot](#)

[Quality format converter \(ASCII-Numeric\)](#)

[Filter by quality](#)

[FASTQ to FASTA converter](#)

[Remove sequencing artifacts](#)

[Barcode Splitter](#)

[Clip adapter sequences](#)

[Collapse sequences](#)

[Draw nucleotides distribution chart](#)

[Compute quality statistics](#)

[Rename sequences](#)

[Reverse-Complement](#)

[Trim sequences](#)

[Combine FASTA and QUAL into FASTQ](#)

[Filter FASTQ reads by quality score and length](#)

[Manipulate FASTQ reads on various attributes](#)

[FASTQ Groomer](#) convert between various FASTQ quality formats

[FASTQ Masker](#) by quality score

[FASTQ joiner](#) on paired end reads

[FASTQ splitter](#) on joined paired end reads

[FASTQ Summary Statistics](#) by column

[FASTQ to FASTA converter](#)

[FASTQ to Tabular converter](#)

[FASTQ Trimmer](#) by column

[FASTQ Quality Trimmer](#) by sliding window

[Tabular to FASTQ converter](#)

[Convert SOLiD output to fastq](#)

[Compute quality statistics](#) for SOLiD data

[Draw quality score boxplot](#) for SOLiD data

NGS: Mapping

[Bowtie2](#) – map reads against reference genome

What: Framework for Tools

The screenshot displays the Galaxy web interface. At the top, there's a navigation bar with 'loud', 'Help', and 'User' menus, and a 'Using 3%' status indicator. The main area is divided into two panels. The left panel shows a workflow with steps like '0.1)' and 'Options'. The right panel, titled 'History', lists a series of datasets and operations: 'Basic Protocol 3: ChIP-Seq' (15 shown, 1.6 GB), '15: Filter pileup on data 14', '14: Generate pileup on data 13: converted pileup', '13: SAM-to-BAM on data 6: converted BAM', '12: MACS on data 5 and data 6 (html report)', '11: MACS on data 5 and data 6 (control: wig)', '10: MACS on data 5 and data 6 (treatment: wig)' (~31,000,000 lines, format: wig, database: mm9). Each entry has icons for viewing, editing, and deleting. At the bottom, there's a section for 'display in IGB View' and 'display at UCSC main'.

3548 valid tools on Aug 31, 2015

Search

- [Search for valid tools](#)
- [Search for workflows](#)

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	53
ChIP-seq		
Combinatorial Selections		
Computational chemistry		
Convert Formats		
Data Managers		
Data Source		
Fasta Manipulation		
Fastq Manipulation		
Genome-Wide Association Study		
Genomic Interval Operations		
Graphics		
Imaging	Utilities to support imaging	
Metabolomics	Tools for use in the study of Metabolomics	
Metagenomics	Tools enabling the study of metagenomes	
Micro-array Analysis	Tools for performing micro-array analysis	
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing	
Ontology Manipulation	Tools for manipulating ontologies	
Phylogenetics	Tools for performing phylogenetic analysis	
Proteomics	Tools enabling the study of proteins	
RNA	Utilities for RNA	
SAM	Tools for manipulating alignments in the SAM format	
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	
Statistics	Tools for generating statistics	
Systems Biology	Systems biology tools	
Text Manipulation	Tools for manipulating data	
Tool Dependency Packages	Repositories that contain third-party tool dependency pack installation definitions	
Tool Generators	Tools that make or help make new tools	

**What: Framework
for Tools**

3548 valid tools on Aug 31, 2015

Search

- [Search for valid tools](#)
- [Search for workflows](#)

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	53
ChIP-seq		
Combinatorial Selections		
Computational chemistry		
Convert Formats		
Data Managers		
Data Source		
Fasta Manipulation		
Fastq Manipulation		
Genome-Wide Association Study		
Genomic Interval Operations		
Graphics		
Imaging	Utilities to support imaging	
Metabolomics	Tools for use in the study of Metabolomics	
Metagenomics	Tools enabling the study of metagenomes	
Micro-array Analysis	Tools for performing micro-array analysis	
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing	
Ontology Manipulation	Tools for manipulating ontologies	
Phylogenetics	Tools for performing phylogenetic analysis	
Proteomics	Tools enabling the study of proteins	
RNA	Utilities for RNA	
SAM	Tools for manipulating alignments in the SAM format	
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	
Statistics	Tools for generating statistics	
Systems Biology	Systems biology tools	
Text Manipulation	Tools for manipulating data	
Tool Dependency Packages	Repositories that contain third-party tool dependency pack installation definitions	
Tool Generators	Tools that make or help make new tools	

What: Framework for Tools


Have mass spec data?
Galaxy-P!

Three ways to Galaxy-P...

Public usegalaxy.org

Local getgalaxy.org

Cloud biocloudcentral.msi.umn.edu

or maybe 4...  [@usegalaxy](https://twitter.com/usegalaxy)

FACE-IT | Galaxy

Tools

Search tools

Get Data
 Create
 Models
 Single-ECM Tools
 Sensitivity Analysis Tools
 Multi-ECM Tools
 ELM-EM
 Agraph

Workflows
 All workflows

FACE-IT is supported by the NSF cyberSEES program award No. 0512702

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

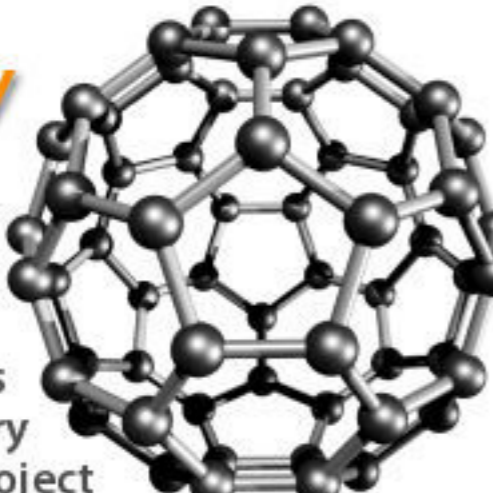
OPEN SOURCE
 DRUG DISCOVERY

OSDDLinux
 LiveGalaxy



ballaxy

Powered by the
 Biochemical
 Algorithms
 Library
 Project



Workflow4metabolomics

VERSION 2.1

LC/MS
 GC/MS
 NMR

MS
 Common



SYM
 WYS

Got symmetry?
 Find out.



Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

Get Data
 Cellular Imaging
 Medical Imaging
 CT Reconstruction
 Copy Dataset to FTP folder
 CT Reconstruction Create a slice from a sinogram
 Center of Rotation Find center of rotation in a sinogram
 Sinogram Creation and Preprocessing Create sinograms from projections

Image filters
 Image Operations
 Image tools

Workflows
 All workflows

Welcome to Cloud-based Image Analysis and Processing Toolbox...

CloudBased
 Image Analysis
 & Processing Toolbox

More information can be found on the NECTAR website, and the project blog.

This project is supported in part by NECTAR, and CSIRO.

Related projects:
 Characterisation Virtual Lab
 Genomics Virtual Lab

History

Unnamed history
 0 bytes

Your history is empty.
 Click 'Get Data' on the left pane to start.

Climate Change

Proteomics
 Metabolomics
 Drug Discovery
 Cosmology
 Image Analysis
 Social Science

CoSSci
 Galaxy for Complex Social Sciences



Natural Language

What: Uniform interface to Tools

Galaxy

Tools

search tools

Get Data

Send Data

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Gro

NGS: QC and manipula

NGS: Mapping

NGS: RNA-seq

NGS: SAMtools

NGS: BAM Tools

NGS: Picard

NGS: VCF Manipulation

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic I

Statistics

Graph/Display Data

Phenotype Association

snpEff

BEDTools

MACS Model-based Analysis of ChIP-Seq (Galaxy Tool Version 1.0.1) Options

Experiment Name

MACS in Galaxy

Paired End Sequencing

Single End

ChIP-Seq Tag File

6: Tags Chr19 SAM

ChIP-Seq Control File

5: Control Chr19 SAM

Effective genome size

1870000000

default: 2.7e+9

Tag size

36

Band width

300

Select the regions with MFOLD high-confidence enrichment ratio against background to build model

Using 3%

ts

3: ChIP-Seq

on dat

leup on

ted pile

M on da

BAM

ta 5 an

eport)

ta 5 an

l: wig)

ata 5 an

ment: wi


es

tabase: mm9

ew

main

What: Automatically records all aspects of analysis

 Galaxy

Analyze Data Workflow Shared Data Visualization Cloud

Tools

search tools

[Get Data](#)
[Send Data](#)
[Lift-Over](#)
[Text Manipulation](#)
[Convert Formats](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[NGS: QC and manipulation](#)
[NGS: Mapping](#)
[NGS: RNA-seq](#)
[NGS: SAMtools](#)
[NGS: BAM Tools](#)
[NGS: Picard](#)
[NGS: VCF Manipulation](#)
[Extract Features](#)
[Fetch Sequences](#)
[Fetch Alignments](#)
[Get Genomic Scores](#)
[Operate on Genomic Intervals](#)
[Statistics](#)
[Graph/Display Data](#)
[Phenotype Association](#)
[snpEff](#)
[BEDTools](#)

MACS Model-based Analysis of ChIP-Seq (Galaxy Tool Version 1.0.1)

Experiment Name
MACS in Galaxy

Paired End Sequencing
Single End

ChIP-Seq Tag File
6: Tags Chr19 SAM

ChIP-Seq Control File
5: Control Chr19 SAM

Effective genome size
1870000000
default: 2.7e+9

Tag size
36

Band width
300

Pvalue cutoff for peak detection
0.00001
default: 1e-5

Select the regions with MFOLD high-confidence enrichment ratio against to build model

History

search datasets

Basic Protocol 3: ChIP-Seq
15 shown
1.6 GB

15: Filter pileup on data 14

14: Generate pileup on data 13: converted pile up

13: SAM-to-BAM on data 6: converted BAM

12: MACS on data 5 and data 6 (html report)

11: MACS on data 5 and data 6 (control: wig)

10: MACS on data 5 and data 6 (treatment: wig)
~31,000,000 lines
format: wig, database: mm9
display in IGB View
display at UCSC main

1

What: History / Analysis management

 **Galaxy**

Analyze DataWorkflowShared DataVisualizationCloud




Tools

search tools

[Get Data](#)
[Send Data](#)
[Lift-Over](#)
[Text Manipulation](#)
[Convert Formats](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[NGS: QC and manipulation](#)
[NGS: Mapping](#)
[NGS: RNA-seq](#)
[NGS: SAMtools](#)
[NGS: BAM Tools](#)
[NGS: Picard](#)
[NGS: VCF Manipulation](#)
[Extract Features](#)
[Fetch Sequences](#)
[Fetch Alignments](#)
[Get Genomic Scores](#)
[Operate on Genomic Intervals](#)
[Statistics](#)
[Graph/Display Data](#)
[Phenotype Association](#)
[snpEff](#)
[BEDTools](#)

MACS Model-based Analysis of ChIP-Seq (Galaxy Tool Version 1.0.1)
Experiment Name
MACS in Galaxy
Paired End Sequencing
Single End
ChIP-Seq Tag File
6: Tags Chr19 SAM
ChIP-Seq Control File
5: Control Chr19 SAM
Effective genome size
1870000000
default: 2.7e+9
Tag size
36
Band width
300
Pvalue cutoff for peak detection
0.00001
default: 1e-5
Select the regions with MFOLD high-confidence enrichment ratio against to build model

History



HISTORY LISTS
Saved Histories
Histories Shared with Me

CURRENT HISTORY
Create New
Copy History
Copy Datasets
Share or Publish
Extract Workflow
Dataset Security
Resume Paused Jobs
Collapse Expanded Datasets
Unhide Hidden Datasets
Delete Hidden Datasets
Purge Deleted Datasets
Show Structure
Export Citations
Export to File
Delete
Delete Permanently

OTHER ACTIONS
Import from File

What: History / Analysis management

Galaxy Analyze Data Workflow Shared Data Visualization Cloud

Tools search tools

- [Get Data](#)
- [Send Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Convert Formats](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: RNA-seq](#)
- [NGS: SAMtools](#)
- [NGS: BAM Tools](#)
- [NGS: Picard](#)
- [NGS: VCF Manipulation](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Phenotype Association](#)
- [snEff](#)
- [BEDTools](#)

MACS Model-based Analysis of ChIP-Seq (Galaxy Tool Version 1.0.1)

Experiment Name
MACS in Galaxy

Paired End Sequencing
Single End

ChIP-Seq Tag File
6: Tags Chr19 SAM

ChIP-Seq Control File
5: Control Chr19 SAM

Effective genome size
1870000000
default: 2.7e+9

Tag size
36

Band width
300

Pvalue cutoff for peak detection
0.00001
default: 1e-5

Select the regions with MFOLD high-confidence enrichment ratio against to build model

History Refresh Settings

HISTORY LISTS

- Saved Histories
- Histories Shared with Me

CURRENT HISTORY

- Create New
- Copy History
- Copy Datasets
- Share or Publish**
- Extract Workflow**
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- Unhide Hidden Datasets
- Delete Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export Citations
- Export to File
- Delete
- Delete Permanently

OTHER ACTIONS

- Import from File

What: Create reusable workflows

Galaxy

Analyze DataWorkflowShared DataVisualizationCloudHelpUser

Workflow Canvas | Workflow constructed from history 'Basic Protocol 3: ChIP-Seq'

Input dataset
output

FASTQ Groomer

File to groom

output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)

Input dataset
output

FASTQ Groomer

File to groom

output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)

Map with Bowtie for Illumina

FASTQ file

output (sam)

output_suppressed_reads_l (fastq)

output_suppressed_reads_r (fastq)

output_unmapped_reads_l (fastq)

output_unmapped_reads_r (fastq)

Map with Bowtie for Illumina

FASTQ file

output (sam)

output_suppressed_reads_l (fastq)

output_suppressed_reads_r (fastq)

output_unmapped_reads_l (fastq)

output_unmapped_reads_r (fastq)

MACS

ChIP-Seq Tag File

ChIP-Seq Control File

output_bed_file (bed)

output_xls_to_interval_peaks_file (interval)

output_xls_to_interval_negative_peaks_file (interval)

output_treatment_wig_file (wig)

output_control_wig_file (wig)

output_extra_files (html)

SAM-to-BAM

SAM File to Convert

output1 (bam)

Filter pileup

Select dataset

out_file1 (tabular, interval)

Generate pileup

Select the BAM file to generate the pileup file for

output1 (tabular)

Generate pileup from BAM

dataset (Galaxy Tool Version 1.1.1)

Will you select a reference genome from your history or use a built-in index?

Use a built-in index

Select the BAM file to generate the pileup file for

Data input 'input1' (bam)

☒ Whether or not to print the mapping quality as the last column

Print the mapping quality as t...

Makes the output easier to parse, but is space inefficient

☒ Whether or not to print only output pileup lines containing indels

Print all lines

☒ Where to cap mapping quality

60

Call consensus according to MAQ model?

Yes

☒ Theta parameter (error dependency coefficient) in the MAQ consensus calling model

0.85

☒ Number of haplotypes in the sample

2

What: Sharing and Publishing

The screenshot displays the Galaxy web interface. The top navigation bar includes the Galaxy logo, 'Analyze Data', 'Workflow', 'Shared Data', 'Data Libraries', 'Cloud', 'Help', and 'User'. The 'Shared Data' dropdown menu is open, showing options: 'Data Libraries', 'Data Libraries Beta', 'Published Histories' (highlighted with a mouse cursor), 'Published Workflows', 'Published Visualizations', and 'Published Pages'. The left sidebar contains a 'Tools' section with a search bar and various tool categories like 'Get Data', 'Send Data', 'Text Manipulation', etc. The main workspace shows the 'MACS Model-based Analysis' tool configuration, including fields for 'Experiment Name', 'Paired End Sequencing', 'ChIP-Seq Tag File', 'ChIP-Seq Control File', 'Effective genome size', 'Tag size', 'Band width', 'Pvalue cutoff for peak detection', and a description. The right sidebar shows the 'History' section with a search bar and a list of recent jobs, including 'Basic Protocol 3: ChIP-Seq' and '15: Filter pileup on data 14'.

Galaxy

Analyze Data Workflow **Shared Data** Data Libraries Cloud Help User Using 3%

Tools

search tools

[Get Data](#)
[Send Data](#)
[Lift-Over](#)
[Text Manipulation](#)
[Convert Formats](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[NGS: QC and manipulation](#)
[NGS: Mapping](#)
[NGS: RNA-seq](#)
[NGS: SAMtools](#)
[NGS: BAM Tools](#)
[NGS: Picard](#)
[NGS: VCF Manipulation](#)
[Extract Features](#)
[Fetch Sequences](#)
[Fetch Alignments](#)
[Get Genomic Scores](#)
[Operate on Genomic Intervals](#)
[Statistics](#)
[Graph/Display Data](#)
[Phenotype Association](#)
[snpEff](#)
[BEDTools](#)

MACS Model-based Analysis

Experiment Name
MACS in Galaxy

Paired End Sequencing
Single End

ChIP-Seq Tag File
6: T

ChIP-Seq Control File
5: C

Effective genome size
1870000000
default: 2.7e+9

Tag size
36

Band width
300

Pvalue cutoff for peak detection
0.00001
default: 1e-5

Select the regions with MFOLD high-confidence enrichment ratio against background to build model

History

search datasets

Basic Protocol 3: ChIP-Seq
15 shown
1.6 GB

15: Filter pileup on data 14

14: Generate pileup on data 13: converted pile up

13: SAM-to-BAM on data 6: converted BAM

12: MACS on data 5 and data 6 (html report)

11: MACS on data 5 and data 6 (control: wig)

10: MACS on data 5 and data 6 (treatment: wig)
~31,000,000 lines
format: wig, database: mm9
display in IGB [View](#)
display at UCSC [main](#)

1

What: Sharing and Publishing



Published Histories


search name, annotation, owner, and tags



[Advanced Search](#)

Name	Annotation	Owner	Community Rating↓	Community Tags
Infravec		dan-lawson	★★★★★	
ChIP-seq shared data		chip-seq-helin-group	★★★★★	chip illumina
Galaxy vs MEGAN	Comparison of Galaxy vs. MEGAN pipeline.	aun1	★★★★★	megan galaxy metagenomics
TRAPLINE: miRNA Targets Input	This history includes the optionally miRNA target prediction files of TRAPLINE. www.sbi.uni-rostock.de/RNAseqTRAPLINE	mwolfien	★★★★★	target prediction mirna
RNA-seq shared data		rna-seq-helin-group	★★★★★	illumina rnaseq
Galaxy Variant 101	Mother-Child mitochondrial variation analysis. See Page https://usegalaxy.org/u/galaxyproject/p/galaxy-101-ngs-variant	galaxyproject	★★★★★	
MOL470 Pset3 All		jbgreisman	★★★★★	
SM_1186088	Datasets correspond to our paper published in Science by Peleg et al. entitled : Altered histone acetylation is associated with age-dependent memory...	publicdata	★★★★★	
SNP Calling		jallen	★★★★★	

What: Publishing *Semantics*



Analyze DataWorkflowShared DataVisualizationCloudHelpUser

Using 3%

Published Pages | mwolfien | TRAPLINE – Manual

TRAPLINE: A standardized and automated pipeline for RNA sequencing data analysis and evaluation





Markus Wolfien, Christian Rimbach, Ulf Schmitz, Julia Jeannine Jung, Stefan Krebs, Gustav Steinhoff, Robert David, and Olaf Wolkenhauer

Department of Systems Biology and Bioinformatics, University of Rostock, 18057 Rostock, Germany
Reference und Translation Center for Cardiac Stem Cell Therapy (RTC), University of Rostock, 18057 Rostock, Germany
Gene Center Munich, LMU Munich, 81377 Munich, Germany
Stellenbosch Institute of Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, 7602 Stellenbosch, South Africa
Correspondence to: markus.wolfien@uni-rostock.de

We critically compare and evaluate state-of-the-art bioinformatics approaches and present a workflow that integrates the best performing data analysis and data evaluation methods in a Transparent, Reproducible and Automated Pipeline (TRAPLINE) for RNA sequencing data analysis. A comparative transcriptomics analysis with TRAPLINE results in a set of differentially expressed genes, their corresponding protein-protein interactions, a analysis of differential splicing and promoter testing and an integrated miRNA target prediction. Ultimately, the user will receive a ready-to-use file which can be imported to Cytoscape.

TRAPLINE supports NGS research by providing a workflow that requires no bioinformatics skills and decreases the processing time of the analysis.


Our pipeline is implemented in the biomedical research platform Galaxy and is freely accessible via:

[Galaxy Workflow | RNAseqTRAPLINE](#)
RNA sequencing data analysis in a Transparent Reproducible and Automated Pipeline – TRAPLINE.

Step by Step instructions for the usage:

- Do your experiments (Illumina, SOLiD, Solexa Sequencing) and obtain the FASTQ files
- Note: the analysis is predefined for the comparison of two experimental conditions with a triplicate for each experimental setup
- Go to the Galaxy website <https://usegalaxy.org>
- If you are new to Galaxy please create an account

About this Page




Author


mwolfien

Related Pages

[All published pages](#)
[Published pages by mwolfien](#)

Rating

Community 
(2 ratings, 5.0 average)


Yours 

Tags

Community:

[workflow](#)[protein_interaction](#)[rnaseq](#)[analysis&evaluation](#)[mirna_prediction](#)

Yours:



Galaxy is available several ways ...

<http://galaxyproject.org>

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

Galaxy

Analyze DataWorkflowShared DataVisualizationCloudHelpUserUsing 3%

Tools

search tools

Get Data

Send Data

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

NGS: QC and manipulation

NGS: Mapping

NGS: RNA-seq

NGS: SAMtools

NGS: BAM Tools

NGS: Picard

NGS: VCF Manipulation

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

snpEff

BEDTools

Genome Diversity

EMBOSS

Galaxy 101

Start small

The very first tutorial you need

Tweets

NIH BD2K @NIH_BD2K1h

Submit #BD2K #Hackathon Proposals to the BD2K Centers Coord. Center! Due OCT15 Read more at ow.ly/SIUkm pic.twitter.com/2bUDJh1tJZ

Retweeted by Galaxy Project

Show Photo

Dawei Lin @iGenomics23h

@mike_schatz My former group at UC Davis has been update an AMLI with Galaxy bioinformatics.ucdavis.edu/software/

Retweeted by Galaxy Project

Expand

Tweet to @galaxyproject

PENNSSTATE

JOHNS HOPKINS UNIVERSITY

TACC

iPlant Collaborative™

A free for everyone web service:

<http://usegalaxy.org>

A free (for everyone) web server integrating a wealth of tools, compute resources, petabytes of reference data and permanent storage




However, *a centralized solution cannot support the different analysis needs of the entire world.*



Explore the
Galaxy with
RNA-Rocket

PATHOGENPORTAL
THE BIOINFORMATICS RESOURCE CENTERS PORTAL

Galaxy / Metabiome Portal



The Microbiome Analysis Center
Life on a Smaller Scale

Welcome to the Metabiome Portal @ GMU

We have developed the MAC Metabiome Portal, a flexible and customizable web browser, with the ability to display, analyze, and visualize microbiome data. The portal is designed to be a central hub for microbiome data, providing a single point of access to a wide range of microbiome data, including raw data, processed data, and analysis results. The portal is designed to be a central hub for microbiome data, providing a single point of access to a wide range of microbiome data, including raw data, processed data, and analysis results.

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

BGI

(GIGA)ⁿ Galaxy
by CBIIT

Integrated publishing of workflows from GIGAⁿ SCIENCE

Cistrome



A Galaxy Server
dedicated to
ChIP-* analysis




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 as Open Source Software

Galaxy is installed in locations around the world.

<http://getgalaxy.org>

Galaxy is available on the Cloud



<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

Why Galaxy *inside Core Facilities?*

Get the advantages of Galaxy in your core

Job tracking

Analysis histories

Reusability

Reproducibility

Data Management

Galaxy API

[Galaxy Code documentation](#) » [lib](#) » [galaxy Package](#) » [webapps Package](#) » [galaxy Package](#) » [previous](#) [next](#) [modules](#) [index](#)

Project Versions

latest

RTD Search

Full-text doc search.

Table Of Contents

- Galaxy API Documentation
 - Background
 - Quickstart
 - API Controllers
 - datasets Module
 - folder_contents Module
 - folders Module
 - forms Module
 - genomes Module
 - group_roles Module
 - group_users Module
 - groups Module
 - histories Module
 - history_contents Module
 - item_tags Module
 - libraries Module
 - library_contents Module
 - permissions Module
 - quotas Module
 - request_types Module
 - requests Module
 - roles Module

Galaxy API Documentation

Background

In addition to being accessible through a web interface, Galaxy can now also be accessed programmatically, through shell scripts and other programs. The web interface is appropriate for things like exploratory analysis, visualization, construction of workflows, and rerunning workflows on new datasets.

The web interface is less suitable for things like

- Connecting a Galaxy instance directly to your sequencer and running workflows whenever data is ready
- Running a workflow against multiple datasets (which can be done with the web interface, but is tedious)
- When the analysis involves complex control, such as looping and branching.

The Galaxy API addresses these and other situations by exposing Galaxy internals through an additional interface, known as an Application Programming Interface, or API.

Quickstart

Log in as your user, navigate to the API Keys page in the User menu, and generate a new API key. Make a note of the API key, and then pull up a terminal. Now we'll use the `display.py` script in your `galaxy/scripts/api` directory for a short example:

```
% ./display.py my_key http://localhost:4096/api/histories
Collection Members
-----
#1: /api/histories/8c49be448cfe29bc
    name: Unnamed history
    id: 8c49be448cfe29bc
#2: /api/histories/33b43b4e7093c91f
    name: output test
    id: 33b43b4e7093c91f
```

<https://galaxy-dist.readthedocs.org/>

Why Galaxy *for Core Facility Clients?*

Empower your clients to actually use the data
you generate for them **without**

learning a programming language,
command line / shell interfaces, Linux
package management, ...

or **extensive hand-holding** from core facility
staff

Empower your clients *with Galaxy*:
Low hanging fruit

Point them at a Galaxy server for their research domain when you give them their data.

bit.ly/gxyServers

Empower your clients with Galaxy:
Moderate

Deliver data inside a Galaxy instance
with appropriate tools and reference datasets,
inside

a virtual machine image,
a Docker container,
or an Amazon Machine Image (AMI)

Empower your clients with Galaxy:
High

Deliver data inside
a core hosted, or institution hosted
Galaxy instance.

Open discussion:

What is the role of cores in supporting client data analysis?

Should this be part of your value proposition?

2016 Galaxy Community Conference (GCC2016)

June 25-29, 2016
Bloomington, Indiana

galaxyproject.org/GCC2016



Join us in beautiful

Bloomington, Indiana

for the 2016 Galaxy
Community Conference
and pre-conference activities!

June 25-29, 2016



Considered one of the five
prettiest campuses in the US,
Indiana University is one of
the major public research
universities in the nation, and
home to the National Center
for Genome Analysis Support.



galaxyproject.org/gcc2016

The Galaxy Team



Enis Afgan



Dannon Baker



Dan Blankenberg



Dave Bouvier



Marten Cech



John Chilton



Dave Clements



Nate Coraor



Carl Eberhard



Jeremy Goecks



Sam Guerler



Jen Jackson



Ross Lazarus



Anton Nekrutenko



Nick Stoler



James Taylor



Nitesh Turaga

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

Acknowledgements

Matt Settles
Ann Norton
Bridget the AV Guru

WACD
ABRF

NIH
Johns Hopkins University
Penn State University
Huck Institute



Thanks