

# Accessible, Transparent, Reproducible Research with Galaxy

---

The Genome Analysis Centre (TGAC)

Norwich, UK

12 April 2013

Dave Clements, Emory University

<http://galaxyproject.org/>

**TGAC**   
The Genome Analysis Centre™

 Galaxy

As science becomes **increasingly dependent on computation:**

- How best to ensure that analysis are **reproducible**?
- How can methods best be made **accessible** to scientists?
- How to facilitate **transparent** communication of analyses?

A crisis in genomics research:  
**reproducibility**

# Key Reproducibility Problems

- **Datasets:** not all available, difficult to access
- **Tools:** inaccessible, hard to record details
- **Publication:** results, data, methods separate

# 50 papers citing bwa

31 provide **no** version and **no** settings

8 lists versions

4 list settings

7 lists versions **and** settings

26 do not provide access to data

*Nekrutenko & Taylor, "Next-generation sequencing data interpretation: enhancing reproducibility and accessibility" Nature Reviews Genetics 13, 667-672 (September 2012)  
doi:10.1038/nrg3305*

# Galaxy: an **accessible** analysis system

**Galaxy** Analyze Data Workflow Shared Data Visualization Cloud Admin Help User Using 158.2 GB

Tools

search tools

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- Phenotype Association
- Genome Diversity
- EMBOSS
- NGS TOOLBOX BETA
- NGS: QC and manipulation
- NGS: Mapping

Additional output created by MACS (MACS\_in\_Galaxy)

**Additional Files:**

- [MACS in Galaxy model.pdf](#)
- [MACS in Galaxy model.r](#)
- [MACS in Galaxy model.r.log](#)
- [MACS in Galaxy negative peaks.xls](#)
- [MACS in Galaxy peaks.xls](#)

**Messages from MACS:**

```
INFO @ Wed, 21 Sep 2011 18:28:58:
# ARGUMENTS LIST:
# name = MACS_in_Galaxy
# format = SAM
# ChIP-seq file = /galaxy/main_database/files/003/013/dataset_3013610.dat
# control file = /galaxy/main_database/files/003/013/dataset_3013609.dat
# effective genome size = 1.87e+09
# tag size = 36
# band width = 300
# model fold = 32
# pvalue cutoff = 1.00e-05
# Ranges for calculating regional lambda are : peak_region,1000,5000,1000
INFO @ Wed, 21 Sep 2011 18:28:58: #1 read tag files...
INFO @ Wed, 21 Sep 2011 18:28:58: #1 read treatment tags...
INFO @ Wed, 21 Sep 2011 18:29:05: #1.2 read input tags...
INFO @ Wed, 21 Sep 2011 18:29:20: #1 Background Redundant rate: 0.02
INFO @ Wed, 21 Sep 2011 18:29:20: #1 finished!
INFO @ Wed, 21 Sep 2011 18:29:20: #2 Build Peak Model...
INFO @ Wed, 21 Sep 2011 18:29:33: #2 number of paired peaks: 16551
INFO @ Wed, 21 Sep 2011 18:29:33: #2 finished!
INFO @ Wed, 21 Sep 2011 18:29:33: #2.2 Generate R script for model : MAC
INFO @ Wed, 21 Sep 2011 18:29:33: #3 Call peaks...
INFO @ Wed, 21 Sep 2011 18:29:33: #3 shift treatment data
INFO @ Wed, 21 Sep 2011 18:29:33: #3 merge +/- strand of treatment data
INFO @ Wed, 21 Sep 2011 18:29:34: #3 save the shifted and merged tag cou
INFO @ Wed, 21 Sep 2011 18:29:34: write to MACS_in_Galaxy_MACS_wiggle/tr
INFO @ Wed, 21 Sep 2011 18:31:04: compress the wiggle file using gzip...
INFO @ Wed, 21 Sep 2011 18:31:20: #3 call peak candidates
INFO @ Wed, 21 Sep 2011 18:31:32: #3 shift control data
INFO @ Wed, 21 Sep 2011 18:31:32: #3 merge +/- strand of control data
INFO @ Wed, 21 Sep 2011 18:31:32: #3 save the shifted and merged tag cou
INFO @ Wed, 21 Sep 2011 18:31:32: write to MACS_in_Galaxy_MACS_wiggle/co
```

**History**

CPB2012 - 1.2 GB

BasicProtocol3 - Calling Peaks for ChIP-seq Data

- 12: [MACS on data 5 and data 6 \(html report\)](#) 3.3 Kb format: html, database: mm9
- 11: [MACS on data 5 and data 6 \(control: wig\)](#)
- 10: [MACS on data 5 and data 6 \(treatment: wig\)](#)
- 9: [MACS on data 5 and data 6 \(negative peaks: interval\)](#)
- 8: [MACS on data 5 and data 6 \(peaks: interval\)](#)
- 7: [CTCF Peaks chr19 BED](#)
- 6: [Tags Chr19 SAM](#)
- 5: [Control Chr19 SAM](#)
- 4: [Tags Chr19 groomed](#)
- 3: [Control Chr19 groomed](#)
- 2: [Tags Chr19 ungroomed](#)

# Integrating existing tools into a uniform framework

The image shows a Galaxy tool interface for a tool named 'Cluster'. On the left, an XML file named 'cluster.xml' is open, showing the tool's configuration. The XML includes a description, command interpreter (python), command, inputs (interval, distance, minregions, returntype), and help text. The main interface on the right is titled 'Cluster' and contains the following elements:

- Cluster intervals of:** A dropdown menu showing '1: UCSC Main on Huma..ne (genome)'.
- max distance between intervals:** A text input field containing '1' with '(bp)' below it.
- min number of intervals per cluster:** A text input field containing '2'.
- Return type:** A dropdown menu showing 'Merge clusters into single intervals'.
- Execute** button.
- TIP:** If your query 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.
- Screenscasts!** See Galaxy Interval Operation Screenscasts (right click to open this link in another window).
- Syntax**
  - **Maximum distance** is greatest distance in base pairs allowed between intervals that will be

The status bar at the bottom indicates 'Line: 87 Column: 8 XML'.

- Defined in terms of an abstract interface (inputs and outputs)
- Designed to be as easy as possible for tool authors, while still allowing rigorous reasoning

# Galaxy analysis interface

The screenshot displays the Galaxy web interface for running a MACS (version 1.0.1) analysis. The interface is divided into three main sections: Tools, the tool configuration panel, and History.

**Tools Panel:** A search bar and a list of tool categories including Get Data, Send Data, ENCODE Tools, Lift-Over, Text Manipulation, Convert Formats, FASTA manipulation, Filter and Sort, Join, Subtract and Group, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, Operate on Genomic Intervals, Statistics, Graph/Display Data, Regional Variation, Multiple regression, Multivariate Analysis, Evolution, Motif Tools, Multiple Alignments, Metagenomic analyses, Phenotype Association, Genome Diversity, EMBOSS, NGS TOOLBOX BETA, NGS: QC and manipulation, NGS: Mapping, NGS: SAM Tools, and NGS: GATK Tools (beta).

**MACS (version 1.0.1) Configuration:**

- 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: 1870000000.0 (default: 2.7e+9)
- Tag size: 36
- Band width: 300
- Pvalue cutoff for peak detection: 1e-05 (default: 1e-5)
- Select the regions with MFOLD high-confidence enrichment ratio against background to build model: 32
- Parse xls files into into distinct interval files:
- Save shifted raw tag count at every bp into a wiggle file: Save
- Extend tag from its middle point to a wigextend size fragment: -1 (Use value less than 0 for default (modeled d))
- Resolution for saving wiggle files: [input field]

**History Panel:** Shows a list of previous analyses. The current analysis is highlighted in green:

- 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)
- 9: MACS on data 5 and data 6 (negative peaks: interval)
- 8: MACS on data 5 and data 6 (peaks: interval)
- 7: CTCF Peaks chr19 BED (720 regions, 1 comments; format: bed, database: mm9)
- 1. Chrom 2. Start 3. End 4. Name

1. Chrom	2. Start	3. End	4. Name
chr19	3204536	3204745	MACS_peak_1
chr19	3208324	3208554	MACS_peak_2
chr19	3218881	3211039	MACS_peak_3
chr19	3291948	3292778	MACS_peak_4
chr19	3320635	3321689	MACS_peak_5

- Consistent tool user interfaces that are automatically generated
- History system facilitates and tracks multistep analyses
- Exact parameters of a step can always be inspected, and easily rerun

# Automatically tracks every step of every analysis

**7: Map with Bowtie for Illumina on data 6 and data 5**   

9,073,928 lines, format: sam,  
database: mm9  
Run this job again 

1. QNAME	2. FLAG	3. I
HWI-EAS269:3:1:1449:913	99	chr
HWI-EAS269:3:1:1449:913	147	chr
HWI-EAS269:3:1:709:832	99	chr
HWI-EAS269:3:1:709:832	147	chr
HWI-EAS269:3:1:1422:1087	99	chr
HWI-EAS269:3:1:1422:1087	147	chr



### Map with Bowtie for Illumina

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

Built-ins were indexed using default options

Select a reference genome:

if your genome of interest is not listed - contact Galaxy team

Is this library mate-paired?:

Forward FASTQ file:

Must have Sanger-scaled quality values with ASCII offset 33

Reverse FASTQ file:

Must have Sanger-scaled quality values with ASCII offset 33

Maximum insert size for valid paired-end alignments (-X):

The upstream/downstream mate orientation for valid paired-end alignment against the forward reference strand (--fr/--rf/--ff):

Bowtie settings to use:

For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:

Bowtie produces SAM with several lines of header information by default

# As well as user-generated metadata and annotation...

History Options

Variant Analysis for Sample E18

Tags:

snp × pileup × bowtie ×

demo × sample:e18 ×

Annotation / Notes:  
Perform a variant analysis with default parameters to identify variants in sample E18 that lie in annotated genes.

10: Variants from sample E18

26,742 regions, format: interval, database: mm9

Info:

Tags:

pileup × sample:e18 ×

snps ×

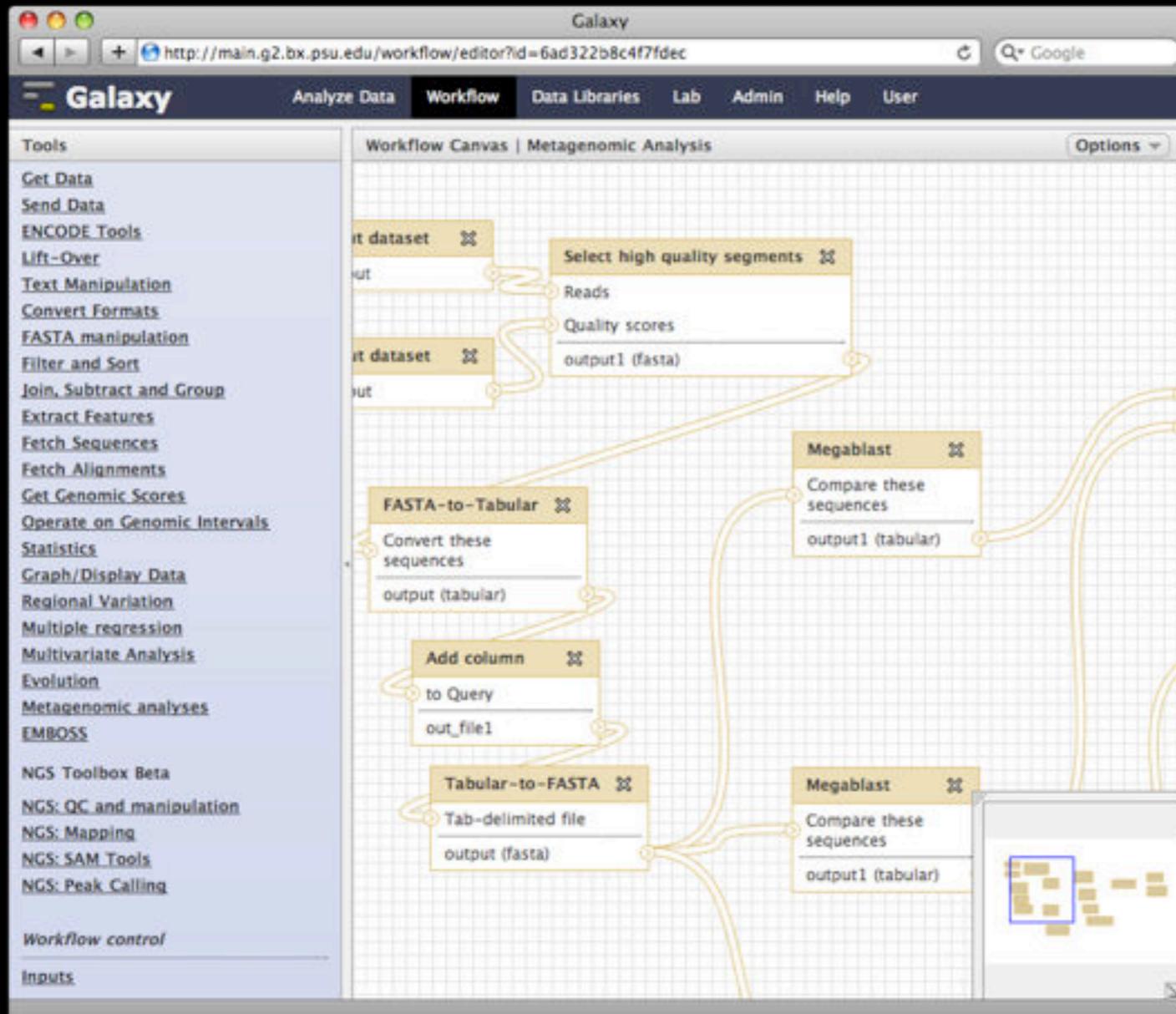
Annotation:

Find variants with coverage  $\geq 30$  and quality score  $\geq 20$ .

| display at UCSC [main](#) | view in [GeneTrack](#) | display at [Ensembl Current](#)

1. Chrom	2. Start	3. End	4	5	6
chr10	6882036	6882037	A	A	107
chr10	14243075	14243076	G	G	96
chr10	14243079	14243080	C	C	106
chr10	14465082	14465083	T	K	173
chr10	14465083	14465084	G	K	144
chr10	14465084	14465085	T	T	117

# Galaxy workflow system



- **Workflows** can be constructed from scratch or extracted from existing analysis histories
- Facilitate reuse, as well as providing precise reproducibility of a complex analysis

# Transparency: Sharing and publishing

The screenshot shows a web browser window displaying a Galaxy page. The browser's address bar shows the URL: <http://main.g2.bx.psu.edu/u/aun1/p/windshield-splatter>. The page title is "Galaxy | Published Page | Windshield Splatter". The main content area features the title "Windshield splatter analysis with the Galaxy metagenomic pipeline: A live supplement" and lists authors: SERGEI KOSAKOVSKY POND<sup>1,2,\*</sup>, SAMIR WADHAWAN<sup>3,6\*</sup>, FRANCESCA CHIAROMONTE<sup>4</sup>, GURUPRASAD ANANDA<sup>1,3</sup>, WEN-YU CHUNG<sup>1,3,7</sup>, JAMES TAYLOR<sup>1,5</sup>, ANTON NEKRUTENKO<sup>1,3</sup> and THE GALAXY TEAM<sup>1\*</sup>. Below the authors, there is a section titled "How to use this document" which explains that the document is a live copy of supplementary materials for a manuscript, providing access to exact analyses and workflows. It includes three interactive elements: a "Galaxy History | Galaxy vs MEGAN" comparison, a "Galaxy History | metagenomic analysis", and a "Galaxy Workflow | metagenomic analysis". Each element has a green plus icon and a share icon. The page also includes a "Supplemental Analysis" section with a link to "Comparison between Galaxy pipeline and Megan".

- All analysis components (datasets, histories, workflows) can be *shared* among Galaxy users and *published*
- Annotation and **Galaxy Pages** allow analyses to be augmented with textual content and provided in the form of an integrated document



## Windshield splatter analysis with the Galaxy metagenomic pipeline

Sergei Kosakovsky Pond<sup>1,2,6,9</sup>, Samir Wadhawan<sup>3,6,7</sup>,  
Francesca Chiaromonte<sup>4</sup>, Guruprasad Ananda<sup>1,3</sup>, Wen-Yu Chung<sup>1,3,8</sup>,  
James Taylor<sup>1,5,9</sup>, Anton Nekrutenko<sup>1,3,9</sup> and The Galaxy Team<sup>1</sup>

### OPEN ACCESS ARTICLE

#### This Article

Published in Advance October 9, 2009, doi: 10.1101/gr.094508.109  
Copyright © 2009 by Cold Spring Harbor Laboratory Press

- » Abstract *Free*
- » Full Text (PDF) *Free*

#### Current Issue

October 2010, 20 (10)



### Footnotes

[Supplemental material is available online at <http://www.genome.org>. All data and tools described in this manuscript can be downloaded or used directly at <http://galaxyproject.org>. Exact analyses and workflows used in this paper are available at <http://usegalaxy.org/u/aun1/p/windshield-splatter>.]

# Windshield splatter analysis with the Galaxy metagenomic pipeline: A live supplement

SERGEI KOSAKOVSKY POND<sup>1,2,\*</sup>, SAMIR WADHAWAN<sup>3,6\*</sup>, FRANCESCA CHIAROMONTE<sup>4</sup>, GURUPRASAD ANANDA<sup>1,3</sup>, WEN-YU CHUNG<sup>1,3,7</sup>, JAMES TAYLOR<sup>1,5</sup>, ANTON NEKRUTENKO<sup>1,3</sup> and THE GALAXY TEAM<sup>1\*</sup>

Correspondence should be addressed to [SKP](#), [JT](#), or [AN](#).

## How to use this document

This document is a live copy of supplementary materials for [the manuscript](#). It provides access to the exact analyses and workflows discussed in the paper, so you can play with them by re-running, changing parameters, or even applying them to your own data. Specifically, we provide the two histories and one workflow found below. You can view these items by clicking on their name to expand them. You can also import these items into your Galaxy workspace and start using them; click on the green plus to import an item. To import workflows you must [create a Galaxy account](#) (unless you already have one) – a hassle-free procedure where you are only asked for a username and password.

This is the Galaxy history detailing the comparison of our pipeline to MEGAN:



[Galaxy History | Galaxy vs MEGAN](#)  
Comparison of Galaxy vs. MEGAN pipeline.



This is the Galaxy history showing a generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and [Figure 3A](#)):



[Galaxy History | metagenomic analysis](#)



This is the Galaxy workflow for generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and [Figure 3B](#)):



[Galaxy Workflow | metagenomic analysis](#)



## Supplemental Analysis

### [Comparison between Galaxy pipeline and Megan](#)

# Give it a spin: [usegalaxy.org/galaxy101](http://usegalaxy.org/galaxy101)

**Galaxy** Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0 bytes

Published Pages | aun1 | Galaxy 101: The first thing you need to try

## Galaxy 101: The first thing you should try

In this very simple example we will introduce you to bare basics of Galaxy:

- Getting data from UCSC
- Performing simple data manipulation
- Understanding Galaxy's History system
- Creating and editing workflows
- Applying workflows to your data

You can watch a step-by-step explanation of this entire tutorial [here](#).

### What are we trying to do?

Suppose you get the following question: "Mom (or Dad) ... Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?". You think to yourself "Wow! This is a simple question ... I know exactly where the data is (at UCSC) but how do I actually compute this?" The truth is, there is really no straightforward way of answering this question in a time frame comparable to the attention span of a 7-year-old. Well ... actually there is and it is called Galaxy. So let's try it...

## 0. Organizing your windows and setting up Galaxy account

### 0.0. Getting your display sorted out

To get the most of this tutorial open two browser windows. One you already have (it is this page). To open the other, right click [this link](#) and choose "Open in a New Window" (or something similar depending on your operating system and browser):

- Open Link in New Window
- Open Link in New Tab
- Download Linked File
- Download Linked File As...
- Add Link to Bookmarks...
- Copy Link

Then organize your windows as something like this (depending on the size of your monitor you may or may not be able to organize things this way, but you get the idea):

### 0.1. Setting up Galaxy account

Go to the **User** link at the top of Galaxy interface and choose **Register** (unless of course you already have an account):

- User
- Login
- Register

Galaxy 101, is a hands-on exercise that demonstrates many Galaxy basics.

Galaxy 101 includes histories, datasets, and workflows, and is itself a *Galaxy Page*.

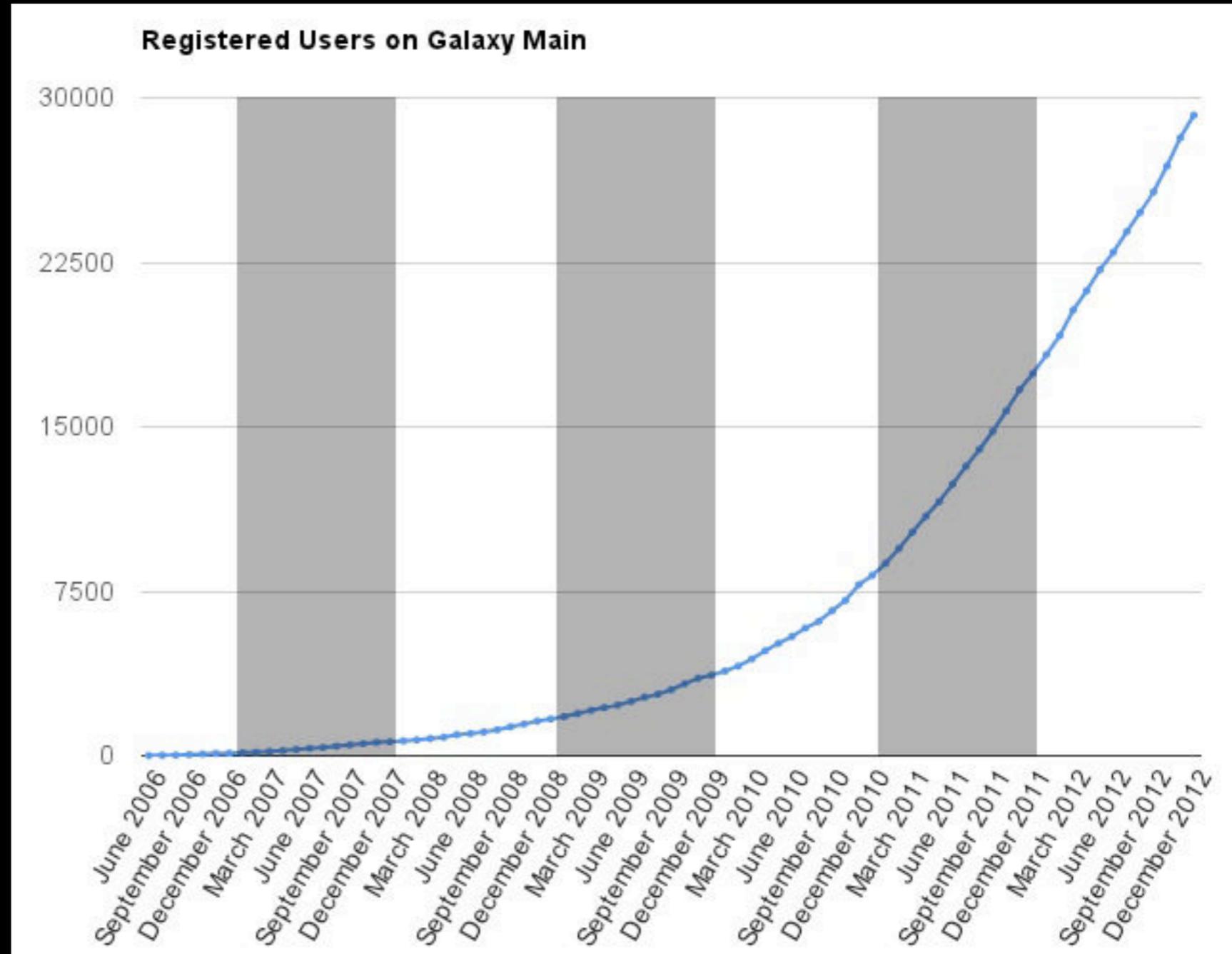
# Galaxy is available ...

- **As a free (for everyone) web service** integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage

<http://usegalaxy.org>

# <http://usegalaxy.org> (a.k.a Main)

- Free public web site
- Anybody can use it
- Persistent
- 150,000+ jobs / month
- Hundreds of tools



<http://bit.ly/gxystats>

# usegalaxy.org: a wealth of tools

## NGS: QC and manipulation

### ILLUMINA DATA

- [FASTQ Groomer](#) convert between various FASTQ qual formats
- [FASTQ splitter](#) on joined paired end reads
- [FASTQ joiner](#) on paired end reads
- [FASTQ Summary Statistics](#) by column

### ROCHE-454 DATA

- [Build base quality distribution](#)
- [Select high quality segments](#)
- [Combine FASTA and QUAL](#) in FASTQ

### AB-SOLID DATA

- [Convert SOLiD output to fastq](#)
- [Compute quality statistics](#) for SOLiD data
- [Draw quality score boxplot](#) for SOLiD data

### GENERIC FASTQ MANIPULATION

- [Filter FASTQ](#) reads by quality score and length
- [FASTQ Trimmer](#) by column
- [FASTQ Quality Trimmer](#) by sliding window
- [FASTQ Masker](#) by quality score

- [Manipulate FASTQ](#) reads on various attributes

- [FASTQ to FASTA](#) converter
- [FASTQ to Tabular](#) converter
- [Tabular to FASTQ](#) converter

### FASTX-TOOLKIT FOR FASTQ DATA

- [Quality format converter](#) (ASCII Numeric)
- [Compute quality statistics](#)
- [Draw quality score boxplot](#)
- [Draw nucleotides distribution chart](#)

- [FASTQ to FASTA](#) converter
- [Filter by quality](#)
- [Remove sequencing artifacts](#)

- [Barcode Splitter](#)
- [Clip adapter sequences](#)
- [Collapse sequences](#)
- [Rename sequences](#)
- [Reverse-Complement](#)

- [Trim sequences](#)

### FASTQ QC

- [FastQC:Read QC](#) reports using FastQC

## NGS: Mapping

### ILLUMINA

- [Map with Bowtie for Illumina](#)

- [Map with BWA for Illumina ROCHE-454](#)

- [Lastz](#) map short reads against reference sequence

- [Megablast](#) compare short reads against htgs, nt, and wgs databases

- [Parse blast XML output](#)

### AB-SOLID

- [Map with Bowtie for SOLiD](#)
- [Map with BWA for SOLiD](#)

## NGS: SAM Tools

- [Filter SAM](#) on bitwise flag values

- [Convert SAM](#) to interval

- [SAM-to-BAM](#) converts SAM format to BAM format

- [BAM-to-SAM](#) converts BAM format to SAM format

- [Merge BAM Files](#) merges BAM files together

- [Generate pileup](#) from BAM dataset

- [Filter pileup](#) on coverage and SNPs

- [Pileup-to-Interval](#) condenses pileup format into ranges of bases

- [flagstat](#) provides simple stats on BAM files

- [rmdup](#) remove PCR duplicates

- [MPileup](#) SNP and indel caller
- [Slice BAM](#) by provided regions

## NGS: GATK Tools (beta)

### ALIGNMENT UTILITIES

- [Depth of Coverage](#) on BAM files
- [Print Reads](#) from BAM files

### REALIGNMENT

- [Realigner Target Creator](#) for use in local realignment
- [Indel Realigner](#) - perform local realignment

### BASE RECALIBRATION

- [Count Covariates](#) on BAM files
- [Table Recalibration](#) on BAM files
- [Analyze Covariates](#) - draw plots

### GENOTYPING

- [Unified Genotyper](#) SNP and indel caller

### ANNOTATION

- [Variant Annotator](#)

### FILTRATION

- [Variant Filtration](#) on VCF files
- [Select Variants](#) from VCF files

### VARIANT QUALITY SCORE RECALIBRATION

- [Variant Recalibrator](#)
- [Apply Variant Recalibration](#)

### VARIANT UTILITIES

- [Validate Variants](#)

- [Eval Variants](#)

- [Combine Variants](#)

## NGS: Indel Analysis

- [Filter Indels](#) for SAM

- [Extract indels](#) from SAM

- [Indel Analysis](#)

## NGS: Peak Calling

- [MACS Model-based Analysis](#) of ChIP-Seq

- [SICER](#) Statistical approach for the Identification of ChIP-Enriched Regions

- [GeneTrack indexer](#) on a BED file

- [Peak predictor](#) on GeneTrack index

## NGS: RNA Analysis

### RNA-SEQ

- [Tophat for Illumina](#) Find splice junctions using RNA-seq data

- [Cufflinks](#) transcript assembly and FPKM (RPKM) estimates for RNA-Seq data

- [Cuffcompare](#) compare assembled transcripts to a reference annotation and track Cufflinks transcripts across multiple experiments

- [Cuffmerge](#) merge together several Cufflinks assemblies

- [Cuffdiff](#) find significant changes in transcript expression

For example, the first 5 pages of NGS tools

# But, it's a big world

Main has lots of tools, storage, processor, users, ...

- But **not all tools** - there are thousands and adding new tools is not taken lightly
- But **not infinite storage and processors** - Main now has job limits and storage quotas

**A centralized solution cannot scale to meet data analysis demands of the whole world**

# Galaxy is available ...

- As a free (for everyone) web service integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- **As open source software** that makes integrating your own tools and data and customizing for your own site simple

<http://getgalaxy.org>

# Local Galaxy Instances

- Galaxy is **designed for local installation and customization**
  - Easily integrate new tools
  - Easy to deploy and manage on nearly any (unix) system
  - Run jobs on existing compute clusters
- Requires an existing computational resource on which to be deployed

**<http://getgalaxy.org>**

# Encourage Local Galaxy Instances

- Support **increasingly decentralized model** and *improve access to existing resources*
- Focus on building **infrastructure to enable the community to integrate and share** tools, workflows, and best practices

# Galaxy Tool Shed

- Allow sites to share “suites” containing tools, datatypes, workflows, sample data, and automated installation scripts for tool dependencies
- Integration with Galaxy instances to automate tool installation and updates

[toolshed.g2.bx.psu.edu](https://toolshed.g2.bx.psu.edu)

# Public Galaxy Servers

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

## Interested in:

ChIP-chip and ChIP-seq?

✓ Cistrome

Statistical Analysis?

✓ Genomic Hyperbrowser

Sequence and tiling arrays?

✓ Oqtans

Text Mining?

✓ DBCLS Galaxy

Reasoning with ontologies?

✓ GO Galaxy

Internally symmetric protein structures?

✓ SymD

# Local Galaxy Instances

- Galaxy is designed for local installation and customization
  - Easily integrate new tools
  - Easy to deploy and manage on nearly any (unix) system
  - Run jobs on existing compute clusters
- Requires an **existing computational resource** on which to be deployed

<http://getgalaxy.org>

# Got your own cluster?

- Galaxy works with DRMAA compliant cluster job schedulers (which is most of them).
- Galaxy is just another client to your scheduler.



# Galaxy is available ...

- As a free (for everyone) web service integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- As open source software that makes integrating your own tools and data and customizing for your own site simple
- **On the Cloud**

<http://usegalaxy.org/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.



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

# Step by Step Instructions on the Wiki for Amazon

## Getting Started with Galaxy CloudMan

This page provides a step-by-step instructions on how to start your own instance of Galaxy on [Amazon Web Services \(AWS\) Elastic Compute Cloud \(EC2\)](#). More general information and instructions about Galaxy CloudMan (GC) can be found [here](#).

**AWS**

[Get Started](#)

[Capacity Planning](#)

[AMIs](#)

[↑ CloudMan](#)

### Contents

1. [Step 1: One Time Amazon Setup](#)
2. [Step 2: Starting a Master Instance](#)
3. [Step 3: Galaxy CloudMan Web Interface](#)
4. [Step 4: Use Galaxy as you normally would](#)
5. [Step 5: Shutting Down](#)

## Step 1: One Time Amazon Setup

1. Because AWS services implement pay-as-you-go access model for compute resources, it is necessary for every user of the service to *register with Amazon*. You will need a credit card to register. (You can apply for a [AWS Education Grant](#) after you register).
2. Once your account has been approved by Amazon (note that this may take up to

### Step 1 Screenshots



# Instant CloudMan

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Cloud', 'Help', and 'User'. A 'Using 0%' indicator is in the top right. The left sidebar contains a 'Tools' panel with a search bar and a list of data sources under 'Get Data'. The main content area displays 'Managing Data' with the text 'Store, Manage, and Share data with Libraries' and 'An in-depth tutorial'. A 'Live Quickies' section is visible below. A tooltip for 'New Cloud Cluster' is shown over the 'Cloud' menu item. The right sidebar shows a 'History' panel with '0 bytes' and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

Launch a CloudMan instance directly from Main

The screenshot shows the 'Launch a Galaxy Cloud Instance' form. It includes the following fields and options:

- Cluster Name:
- Password:
- Key ID:
- Secret Key:
- Instance Share String (optional):
- Instance Type:

Requesting the instance may take a moment, please be patient. Do not refresh your browser or navigate away from the page

# Galaxy Community & Resources: Mailing Lists

## Galaxy-Announce

Project announcements, low volume, moderated

Low volume ( 42 posts, 1600 members in 2012)

## Galaxy-User

Questions about using Galaxy and usegalaxy.org

High volume (2900 posts, 2700 members in 2012)

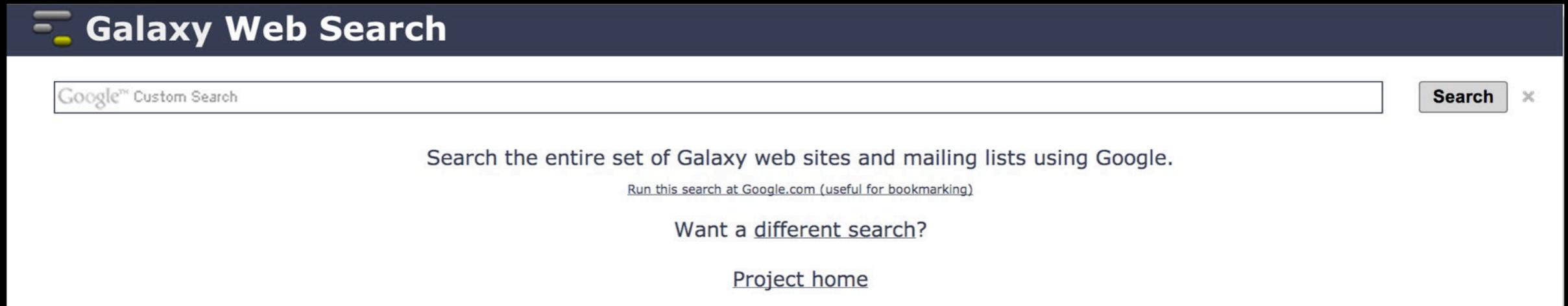
## Galaxy-Dev

Questions about developing for and deploying Galaxy

High volume (4500 posts, 850 members in 2012)

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

# Galaxy Search: <http://galaxyproject.org/search>



Galaxy Web Search

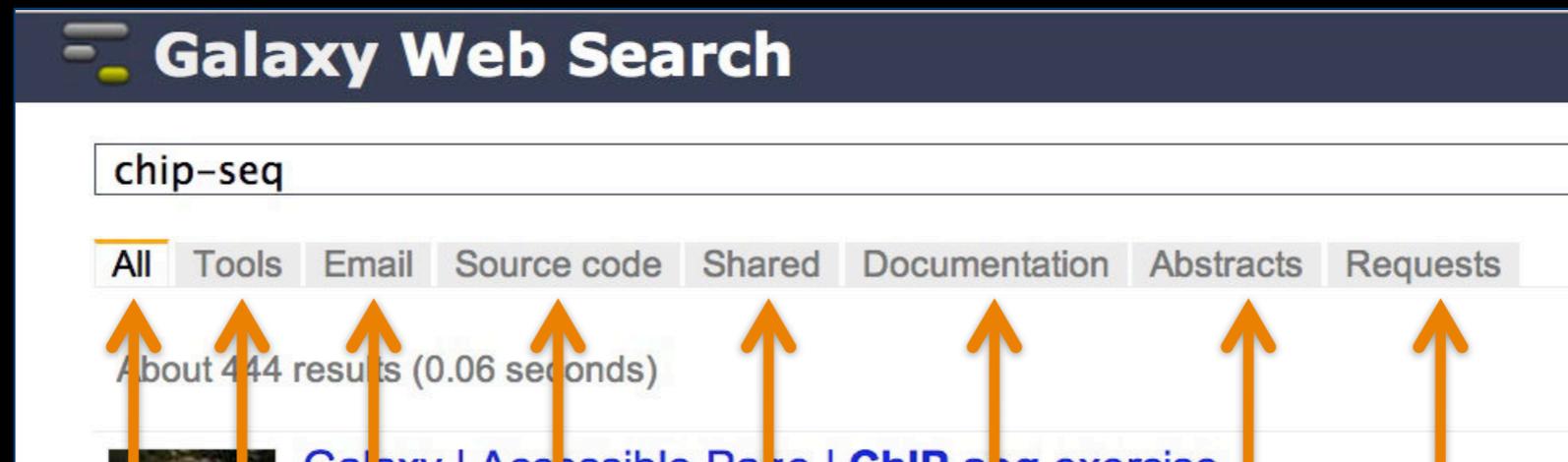
Google™ Custom Search  Search x

Search the entire set of Galaxy web sites and mailing lists using Google.

[Run this search at Google.com \(useful for bookmarking\)](#)

Want a [different search?](#)

[Project home](#)



Galaxy Web Search

chip-seq

All Tools Email Source code Shared Documentation Abstracts Requests

About 444 results (0.06 seconds)

Galaxy | Accessible Page | ChIP-seq exercise

## Find

Everything on ...

Tools for ...

Email about ...

Source code for ...

Published Histories, Pages, Workflows, about ...

Documentation on ...

Papers using Galaxy for ...

Related feature requests

[galaxyproject.org/GCC2013](http://galaxyproject.org/GCC2013)

**Galaxy**  
Community  
Conference

30 June  
- 2 July

2013

OSLO

UiO : University of Oslo

(GIGA)<sup>n</sup>  
SCIENCE

GCC2013  
Training  
Day

STARTING @

€95

galaxyproject.org/GCC2013



The logo for the Galaxy Community Conference 2013 features the text "Galaxy Community Conference" in blue and red, "OSLO" in white on a blue wavy background, and "2013" in large white numbers on a blue background. It also includes the dates "30 June - 2 July" and the University of Oslo logo and name "UiO : University of Oslo". A green map of Norway is overlaid on the right side.

Galaxy  
Community  
Conference

OSLO

30 June  
- 2 July

2013

UiO : University of Oslo

Talk abstracts due **TODAY**



The GIGA Science logo consists of the text "(GIGA)<sup>n</sup> SCIENCE" with a green DNA double helix icon.

(GIGA)<sup>n</sup>  
SCIENCE



The GCC2013 Training Day logo features the text "GCC2013 Training Day" and a blue silhouette of a skier on a red slope.

GCC2013  
Training  
Day

STARTING

@

€95

**Galaxy Wiki** DaveClements Settings Logout | Search:  Titles Text

FrontPage Edit History Actions



**Galaxy** is an open, web-based platform for *accessible, reproducible, and transparent* computational biomedical research.

- **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

This is the Galaxy Community Wiki. It describes all things Galaxy.

### Use Galaxy

Galaxy's [public service web site](#) makes analysis tools, genomic data, tutorial demonstrations, persistent workspaces, and publication services available to any scientist. Extensive [user documentation](#) (applicable to any [public](#) or local Galaxy instance) is available on [this wiki](#) and elsewhere.



### Deploy Galaxy

Galaxy is open source for all organizations. Local Galaxy servers can be set up by [downloading and customizing](#) the Galaxy application.

- [Admin](#)
- [Cloud](#)



### Community & Project

Galaxy has a large and active user community and many ways to [Get Involved](#).

- [Community](#)
- [News](#)
- [Events](#)
- [Support](#)
- [Galaxy Project](#)

### Contribute

- **Users:** Share your histories, workflows, visualizations, data libraries, and [Galaxy Pages](#), enabling others to use and learn from them.
- **Deployers and Developers:** Contribute tool definitions to the [Galaxy Tool Shed](#) (making it easy for others to use those tools on their installations), and code to the core release.
- **Everyone: [Get Involved!](#)**



Topic voting now open!



### Use Galaxy

[Project Server](#) (*Use it!*)  
[Other Servers](#) • [Learn Share](#) • [Search](#)

### Communication

[Support](#) • [News](#)   
[Events](#) • [Twitter](#)  
[Mailing Lists](#) ([search](#))

### Deploy Galaxy

[Get Galaxy](#) • [Cloud Admin](#) • [Tool Config](#)  
[Tool Shed](#) • [Search](#)

### Contribute

[Tool Shed](#) • [Share Issues & Requests](#)  
[Support](#)

### Galaxy Project

[Home](#) • [About Community](#)  
[Big Picture](#)

# Events

# News

**Galaxy Wiki** Dave Clements Settings Log out

Events

## Galaxy Event Horizon

Events with Galaxy-related content are listed here.

Also see the [Galaxy Events Google Calendar](#) for a listing of events and deadlines that are relevant to the Galaxy Community. This is also available as an [RSS feed](#).

If you know of any event that should be added to this page and/or to the Galaxy Event Calendar, please add it here or send it to [outreach@galaxyproject.org](mailto:outreach@galaxyproject.org).

### Upcoming Events





Date	Topic/Event	Venue/Location
April 7-10	GO Galaxy Workshop	Biocuration 2013, Cambridge, United Kingdom
April 7-8	BOSC/Broad Interoperability Hackathon	Cambridge, Massachusetts, United States
April 9-11	Workshop: <i>Integrated Research Data Management for Next Gen Sequencing Analysis Using Galaxy and Globus Online Software-as-a-Service</i>	BioIT World, Boston, Massachusetts, United States
	Talk: <i>Integrated Research Data management and Analysis in NGS using Globus Online, Galaxy and Amazon Web Services</i>	
April 10	Introduction to Galaxy Boot Camp <b>Registration is full</b>	UC Davis Bioinformatics Core Davis, California, United States
April 11	Introduction to RNASeq Boot Camp <b>Registration is full</b>	
April 11	Introduction to Galaxy Workshop	The Genome Analysis Centre, Norwich, United Kingdom
April 12	Next generation sequencing data interpretation: enhancing reproducibility and accessibility	Reed College, Portland, Oregon, United States
May 14-16	Tutorial: <i>Exploring and Enabling Biomedical Data Analysis with Galaxy</i>	Great Lakes Bioinformatics Conference (GLBIO) 2013, Pittsburgh, Pennsylvania, United States
May 16-17	Galaxy Workflows for Bioinformatics Analysis, and Workshop 1A - Galaxy Workflows for Bioinformatics Analysis	Workshop in Next-Generation Sequence Analysis and Metabolomics (WINGS), UNC-Charlotte, North Carolina, United States
May 21	Initiation à l'utilisation de Galaxy	Cycle "Bioinformatique par la pratique" 2013, INRA Jouy-en-Josas, France
May 29	<b>Les deux ateliers sont maintenant complets</b>	
May 22	Analyse de données issues de séquenceurs nouvelle génération sous Galaxy	

## News

Announcements of interest to the Galaxy Community. These can include items from the Galaxy Team or the Galaxy community and can address anything that is of wide interest to the community.

The Galaxy News is also available as an [RSS feed](#).

See [Add a News Item](#) below for how to get an item on this page, and the RSS feed. Older news items are available in the [Galaxy News Archive](#).

### See also

- Distribution News Briefs
- Galaxy Updates
- Galaxy on Twitter
- Events
- Learn
- Support
- About the Galaxy Project

## News Items

February 2013 Galaxy Update

The February 2013 Galaxy Update is now available.

### Highlights:

- Three new public Galaxy servers
- New papers
- Open Positions at five different institutions
- GCC2013 Training Day Topic voting, Registration, and Sponsorships
- January GalaxyAdmins Web Meetup slides and screencast
- Other Upcoming Events and Deadlines
- Galaxy Distributions
- Tool Shed Contributions
- Other News

If you have anything you would like to see in the March *Galaxy Update*, please let us know.

Dave Clements and the Galaxy Team

*Posted to the Galaxy News on 2013-02-01*

GCC2013 Training Day Topics: Vote!

A list of possible topics for the GCC2013 Training Day is now available. Please take a few minutes to review these possibilities and then vote for your favorite three topics.\*

Your votes will determine not only the topics that are offered, but also which topics should be offered more than once, assigned to which rooms, and which ones should not be scheduled at the same time. Your vote matters.

### News Items

- February 2013 Galaxy Update
- GCC2013 Training Day Topics: Vote!
- Galaxy Project Openings
- Jan 11, 2013 Distribution & News Brief
- January 2013 GalaxyAdmins
- January 2013 Galaxy Update
- Dec 20, 2012 Distribution & News Brief
- Galaxy Internships @ EMBL
- Nominate GCC2013 Training Topics
- Dec 3, 2012 Distribution & News Brief
- December 2012 Galaxy Update
- Nov 14, 2012 Distribution & News Brief
- NGS Analysis by Viz. with Trackster
- November 2012 GalaxyAdmins

[News Archive](#)





# Visualization

Send data results to **external** genome browsers:

UCSC, Ensembl, GBrowse, IGV

**Trackster:** Galaxy's genome browser

# Trackster

## View your data from within Galaxy

- ✦ No data transfers to external site
- ✦ Use it locally, even without internet access

## Supports common filetypes

- ✦ BAM, BED, GFF/GTF, WIG

## Unique features

- ✦ custom genomes
- ✦ highly interactive

Published Visualizations | jeremy | GCC2011-1: Viewing and

chr19

625,719 - 682,581

630,000 640,000 650,000 660,000 670,000 680,000

UCSC Main on Human: knownGene (chr19)

Auto (Squish)

UCSC Main on Human: all\_est (chr19)

Dense

UCSC Main on Human: phyloP46wayPrimates (chr19)

Histogram

1

-1

h1-hESC Tophat Mapped Reads

Auto (Squish)

630,000 640,000 650,000 660,000 670,000 680,000

Published Visualizations | jeremy | GCC2011-1: Viewing and chr19 663,032 - 663,110

g e e e g g g e e T C A C C G G C A G G C G C G G G R C G A T C T C C A C G G A G C A G C A G T G G C A G A R G T A C C G T C C G G G A T G C G G C G A C

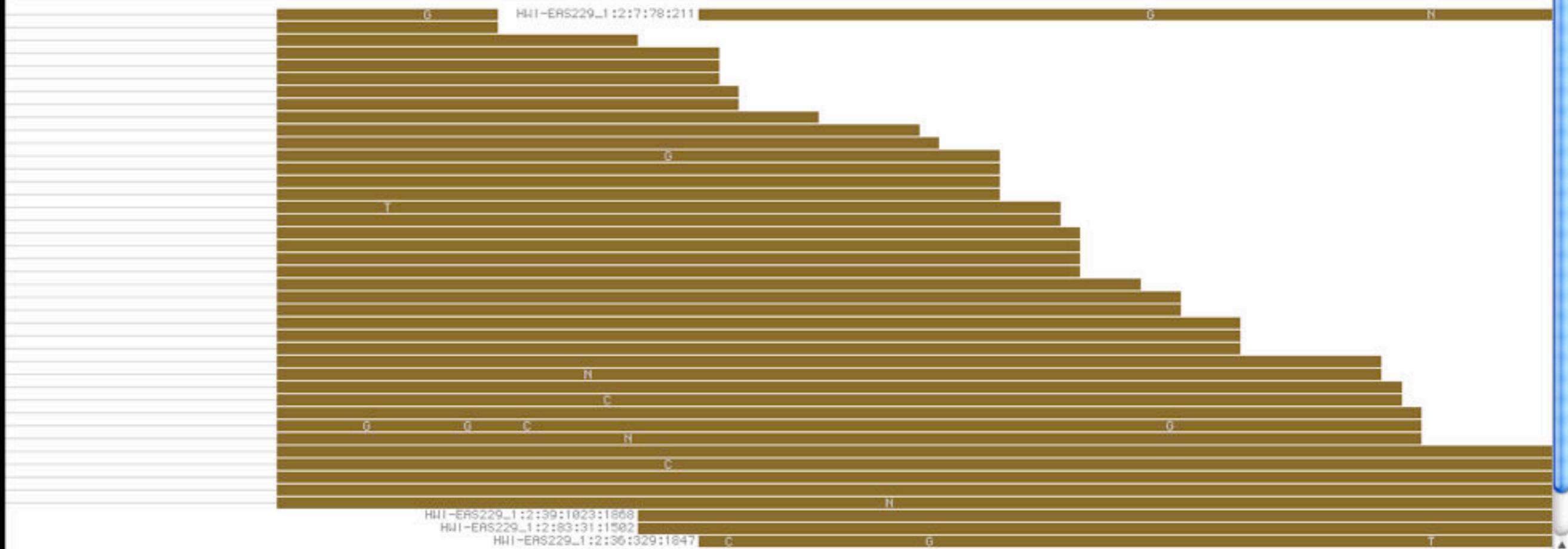
UCSC Main on Human: knownGene (chr19) Auto (Pack)

UCSC Main on Human: all\_est (chr19) Dense

UCSC Main on Human: phyloP46wayPrimates (chr19) Histogram



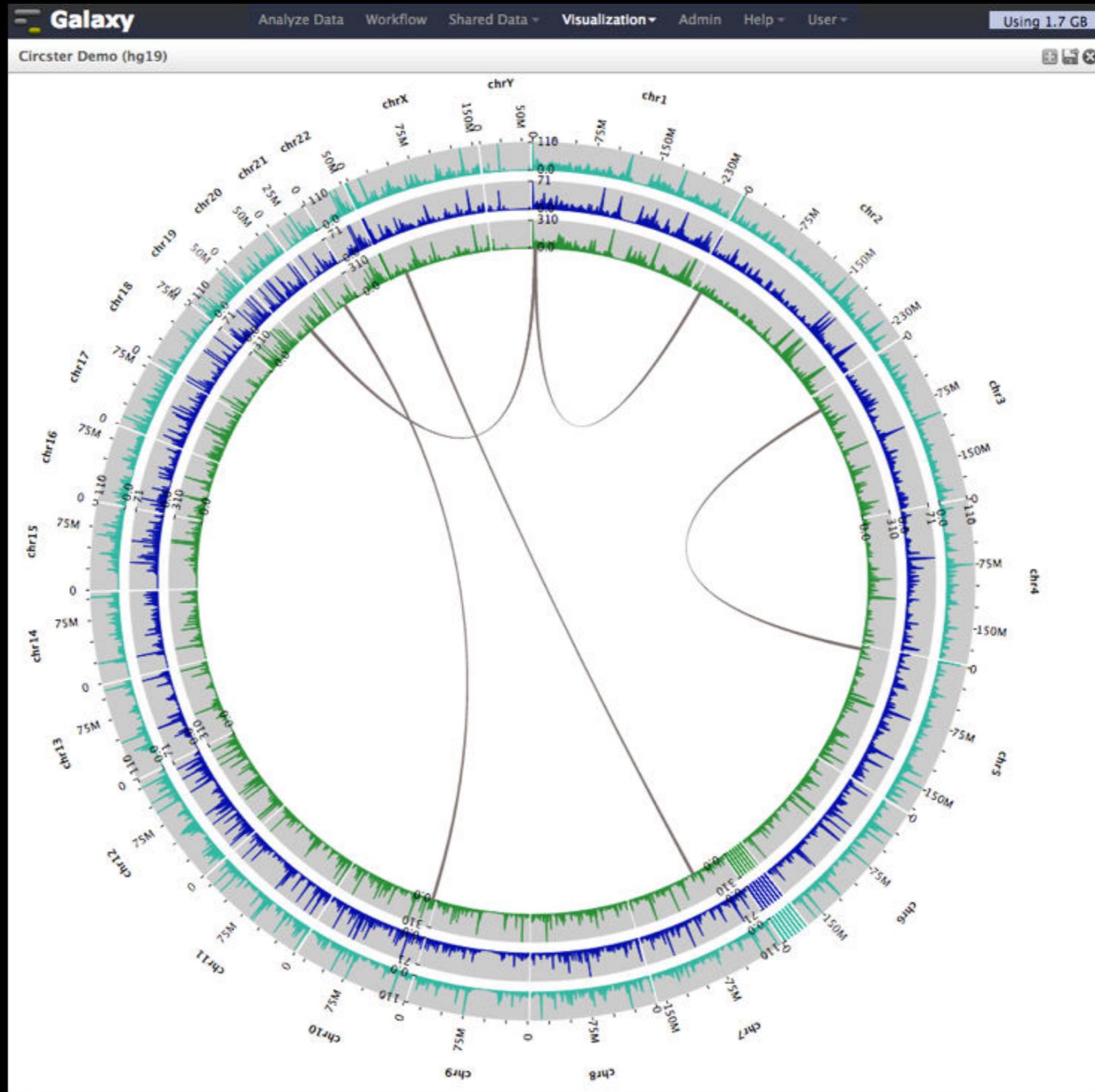
h1-hESC Tophat Mapped Reads Auto (Pack)



h1-hESC Cufflinks assembled transcripts Auto (Pack)

663,040 663,050 663,060 663,070 663,080 663,090 663,100

# Visualization: Circster



Circster: Circos style visualizations

# But really, why *another* genome browser

From static browsing to **visual analysis**

**Visual feedback and experimentation** needed for complex tools with many parameters

**Leverage Galaxy strengths:** a very sound model for abstracting interfaces to analysis tools and already integrates an enormous number

# Dynamic Filtering





# Exploring Parameter Space with Trackster

Galaxy Analyze Data Workflow Shared Data Visualization Cloud

Published items | jeremy | Trackster Demo 2 chr19 1,549,354 - 1,691,104 1,600,000

GM12878 Cufflinks assembled transcripts BEST

h1-hESC Cufflinks assembled transcripts BEST

Tool parameter space visualization

Galaxy Analyze Data Workflow Shared Data Visualization Cloud Help User

### Cufflinks (version 0.0.5)

**Max Intron Length:** 200000 - 400000 samples: 3

**Min Isoform Fraction:** 0.1 - 0.2 samples: 3

**Pre mRNA Fraction:** 0.15

**Perform quartile normalization:** No, Yes

**Use multi-read correct:** No

Execute

### Getting Started

1. Create a parameter tree by using the icons next to the tool's parameter names to add or remove parameters.
2. Adjust the tree by using parameter inputs to select min, max, and number of samples
3. Run the tool with different settings by clicking on tree nodes

Max Intron Length Perform quartile normalMin Isoform Fractl

Root 200000 300000 400000

No Yes No Yes No Yes

0.1 0.15 0.2 0.1 0.15 0.2 0.1 0.15 0.2 0.1 0.15 0.2 0.1 0.15 0.2

chr19:1549354-1691104

Building a new browser for  
your own genome

u.edu/root

flow Shared Data Lab Visualization Cloud Admin Help User Using 30.8 Gb

## Explore the Galaxy!



**Galaxy**  
Community Conference  
2012

July 25-27  
**UIC Forum**  
University of Illinois at Chicago  
<http://galaxyproject.org/GCC2012>

**UIC** 

**This July**  
(For as little as \$135)

### Live Quickies

Advanced fastQ manipulation:  
Galactic quickie # 14

454 Mapping: Single End  
Galactic quickie # 15

Uploading using  
Galactic quickie # 16

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 [BX](#) at [Penn State](#), and the [Biology](#) and [Mathematics and Computer Science](#) departments at [Emory University](#). The [Galaxy Project](#) is supported in part by [NSF](#), [NHGRI](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience](#) at Penn

### History

Unnamed history	23.3 Mb
<b>4: Promoter scan</b>	
<b>3: CRISPR locations</b>	
<b>2: RefSeq</b>	
<b>1: pyrAer.fa</b>	

Promoter potential (WIG)  
 CRISPR elements (BED)  
 Genes (BED)  
**Genome (FASTA)**

**Current Custom Builds:**  
You currently have no custom builds.  
[Show loaded, system-installed builds](#)

**Add a Custom Build**

New Build

Name (eg: Hamster):

Pyrobaculum aeroph

Key (eg: hamster\_v1):

pyroAero

Definition:

**FASTA** [Len File](#) [Len Entry](#)

1: pyrAer.fa

Submit

**FASTA format**

This is a multi-fasta file from your current history that provides the genome sequences for each chromosome/contig in your build.

Here is a snippet from an example multi-fasta file:

```
>chr1
ATTATATATAAGACCACAGAGAGAAATATTTTGCCCGG...
>chr2
GGCGGCCGCGGCGATATAGAACTACTCATTATATATA...
...
```

**History**

Unnamed history	23.3 Mb
<b>4: Promoter scan</b>	
<b>3: CRISPR locations</b>	
<b>2: RefSeq</b>	
<b>1: pyrAer.fa</b>	

### Current Custom Builds:

Name	Key	Number of chrams/contigs	
Pyrobaculum aerophilum str. IM2	pyroAero	1	<button>Delete</button>

[Show loaded, system-installed builds](#)

### Add a Custom Build

#### New Build

Name (eg: Hamster):

Key (eg: hamster\_v1):

Definition:  
[FASTA](#) [Len File](#) [Len Entry](#)

1: pyrAer.fa ↕

Submit

**FASTA format**

This is a multi-fasta file from your current history that provides the genome sequences for each chromosome/contig in your build.

Here is a snippet from an example multi-fasta file:

```
>chr1
ATTATATATAAGACCACAGAGAGAAATATTTTGCCCGG...
>chr2
GGCGGCCGCGGCGATATAGAACTACTCATTATATATA...
...
```

### History ⚙️

Unnamed history <span>📎 📁</span>	23.3 Mb
<b>4: Promoter scan</b> <span>👁️ 0 ✖️</span>	
<b>3: CRISPR locations</b> <span>👁️ 0 ✖️</span>	
<b>2: RefSeq</b> <span>👁️ 0 ✖️</span>	
<b>1: pyrAer.fa</b> <span>👁️ 0 ✖️</span>	

u.edu/root

flow Shared Data Lab Visualization Cloud Admin Help User Using 30.8 Gb

## Software Engineer? Post-Doc?

# The Galaxy Project Wants You!



Live Quickies

- Uploading Data using FTP Galactic quickie # 17
- Managing account histories Galactic quickie # 19

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 [BX](#) at [Penn State](#), and the [Biology](#) and [Mathematics and Computer Science](#) departments at [Emory University](#). The [Galaxy Project](#) is supported in part by [NSF](#), [NHGRI](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience](#) at Penn

### History

- Unnamed history 23.3 Mb
- 4: Promoter scan
- 3: CRISPR locations
- 2: RefSeg
  - 2,603 regions
  - format: bed, database: pyroAero
  - view in [GeneTrack](#)
  - display in IGB [Local](#) [Web](#)
- 1: pyrAer.fa

1. Chrom	2. Start	3. End	4. Name	5	6. Str
chr	0	957	PAE0001	1	+
chr	953	1427	PAE0002	2	-
chr	1470	2397	PAE0005	3	-
chr	2639	3797	PAE0007	4	+
chr	3783	4143	PAE0008	5	+
chr	4315	4456	PAE0009	6	-

Custom build

Visualize in trackster

Galaxy [https://main.g2.bx.psu.edu/tracks?default\\_dbkey=pyroAero&dataset\\_id=f6365eb5c96fc6d...](https://main.g2.bx.psu.edu/tracks?default_dbkey=pyroAero&dataset_id=f6365eb5c96fc6d...) Using 30.8 Gb

Demo (pyroAero) chr 0 - 2,222,429

0 1,000,000 2,000,000

|||RefSeq

0 1,000,000 2,000,000

The image shows a web browser window displaying the Galaxy bioinformatics platform. The browser's address bar shows the URL: [https://main.g2.bx.psu.edu/tracks?default\\_dbkey=pyroAero&dataset\\_id=f6365eb5c96fc6d...](https://main.g2.bx.psu.edu/tracks?default_dbkey=pyroAero&dataset_id=f6365eb5c96fc6d...). The Galaxy interface includes a navigation menu with options like 'Analyze Data', 'Workflow', 'Shared Data', 'Lab', 'Visualization', 'Cloud', 'Admin', 'Help', and 'User'. A status bar at the top right indicates 'Using 30.8 Gb'. The main content area is titled 'Demo (pyroAero)' and shows a genomic track for chromosome 'chr' with a range of '0 - 2,222,429'. A scale bar at the top of the track shows positions '0', '1,000,000', and '2,000,000'. The track itself is labeled '|||RefSeq' and displays a dense series of horizontal lines representing RefSeq annotations, with some lines colored in green and red. A second scale bar is visible at the bottom of the track area, also showing '0', '1,000,000', and '2,000,000'.

Galaxy [https://main.g2.bx.psu.edu/tracks?default\\_dbkey=pyroAero&dataset\\_id=f6365eb5c96fc6d...](https://main.g2.bx.psu.edu/tracks?default_dbkey=pyroAero&dataset_id=f6365eb5c96fc6d...) Using 30.8 Gb

Demo (pyroAero) chr 957,612 - 1,029,613

960,000 970,000 980,000 990,000 1,000,000 1,010,000 1,020,000

||RefSeq

PAE1646	PAE1653	PAE1658	PAE1670	PAE1677	PAE1687	PAE1694	PAE1704	PAE1715	PAE1719	PAE1727	PAE1736	PAE1749	PAE1759
PAE1647	PAE1656	PAE1665	PAE1672	PAE1683	PAE1690	PAE1702	PAE1712	PAE1718	PAE1726	PAE1735	PAE1744	PAE1755	PAE1766
PAE1638	PAE1649	PAE1657	PAE1671	PAE1685	PAE1693	PAE1703	PAE1714	PAE1720	PAE1729	PAE1738	PAE1752	PAE1760	PAE1768
PAE1639	PAE1650	PAE1659	PAE1673	PAE1688	PAE1696	PAE1708	PAE1717a	PAE1723	PAE1731	PAE1740	PAE1756	PAE1764	PAE1772
PAE1641	PAE1651	PAE1662	PAE1674	PAE1689	PAE1698	PAE1709	PAE1722	PAE1730	PAE1742	PAE1761	PAE1769	PAE1777	PAE1785
PAE1642	PAE1652	PAE1663	PAE1676	PAE1695	PAE1710	PAE1725	PAE1734	PAE1748	PAE1757	PAE1765	PAE1773	PAE1781	PAE1789
PAE1643	PAE1667	PAE1678	PAE1699	PAE1711	PAE1724	PAE1733	PAE1741	PAE1753	PAE1762	PAE1770	PAE1778	PAE1786	PAE1794
PAE1645	PAE1668	PAE1679	PAE1700	PAE1713	PAE1721	PAE1728	PAE1737	PAE1745	PAE1754	PAE1763	PAE1771	PAE1779	PAE1787
	PAE1669	PAE1681	PAE1701	PAE1714	PAE1722	PAE1731	PAE1739	PAE1747	PAE1755	PAE1764	PAE1772	PAE1780	PAE1788

960,000 970,000 980,000 990,000 1,000,000 1,010,000 1,020,000

### Select datasets for new tracks

Histories Data Libraries

#### History 'Unnamed history'

search name and filetype

<input type="checkbox"/>	Id	Name	Filetype	Dbkey
<input checked="" type="checkbox"/>	4	Promoter scan	wig	pyroAero
<input checked="" type="checkbox"/>	3	CRISPR locations	bed	pyroAero
<input type="checkbox"/>	2	RefSeq	bed	pyroAero

For 2 selected datasets:

Cancel Insert



Galaxy <https://main.g2.bx.psu.edu/tracks/browser?id=d518ad8573371ac2#chr:21663-122199> Using 30.8 Gb

Demo (james:pyroAero) chr 21,663 - 122,199 100,000

||RefSeq

PAE0055 PAE0064 PAE0075 PAE0087 PAE0100 PAE0111 PAE0120 PAE0131 PAE0148 PAE0165 PAE0178 PAE0198 PAE0208 PAE0218 PAE0226  
PAE0056 PAE0064a PAE0077 PAE0089 PAE0103 PAE0114 PAE0121 PAE0134 PAE0150 PAE0166 PAE0180 PAE0199 PAE0209 PAE0219 PAE0227  
PAE0057 PAE0065 PAE0078 PAE0090 PAE0105 PAE0115 PAE0122 PAE0137 PAE0152 PAE0168 PAE0181 PAE0200 PAE0210 PAE0220 PAE0228  
PAE0061 PAE0068 PAE0082 PAE0096 PAE0109 PAE0118 PAE0128 PAE0146 PAE0164a PAE0201 PAE0211 PAE0221 PAE0229  
PAE0062 PAE0069 PAE0084 PAE0098 PAE0112 PAE0124 PAE0138 PAE0154 PAE0172 PAE0202 PAE0212 PAE0222 PAE0230  
PAE0063 PAE0073 PAE0086 PAE0101 PAE0116 PAE0126 PAE0139 PAE0157 PAE0173 PAE0204 PAE0213 PAE0224 PAE0231  
PAE0051 PAE0067 PAE0079 PAE0091 PAE0106 PAE0117 PAE0129 PAE0151 PAE0170 PAE0205 PAE0214 PAE0225 PAE0232  
PAE0047 PAE0070 PAE0088 PAE0108 PAE0119 PAE0133 PAE0153 PAE0171 PAE0207 PAE0215 PAE0226 PAE0233  
PAE0049 PAE0080 PAE0097 PAE0135 PAE0158 PAE0175 PAE0208 PAE0216 PAE0227 PAE0234  
PAE0049a PAE0081 PAE0141 PAE0160 PAE0176 PAE0209 PAE0217 PAE0228 PAE0235  
PAE0142 PAE0163 PAE0177 PAE0210 PAE0218 PAE0229 PAE0236  
PAE0143 PAE0164 PAE0178 PAE0211 PAE0220 PAE0231 PAE0237  
PAE0144 PAE0167 PAE0179 PAE0212 PAE0221 PAE0232 PAE0238

||Promoter scan



||CRISPR locations

crispr1 crispr2 crispr3

100,000

# Galaxy URLs to Remember

<http://galaxyproject.org>

<http://usegalaxy.org>

<http://getgalaxy.org>

# The Galaxy Team



Enis Afgan



Dannon Baker



Dan Blankenberg



Dave Bouvier



Dave Clements



Nate Coraor



Carl Eberhard



Dorine Francheteau



Jeremy Goecks



Sam Guerler



Jen Jackson



Greg von Kuster



Ross Lazarus



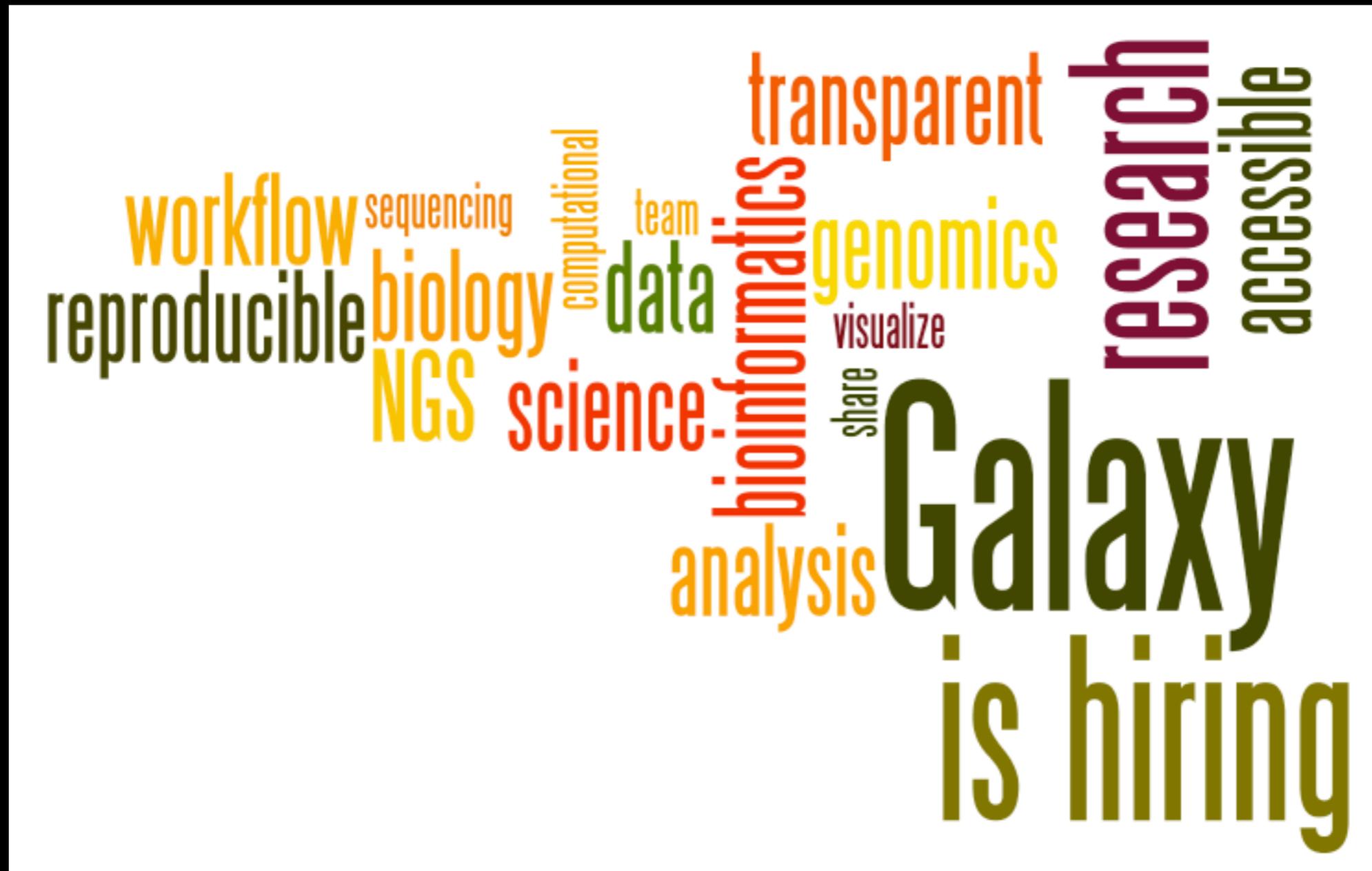
Anton Nekrutenko



James Taylor

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

Galaxy is hiring post-docs and software engineers  
at both Emory and Penn State.



Please help.

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

# Acknowledgements

Vicky Schneider-Gricar  
Helen Tunney  
Robert Davey

The Galaxy Team  
You!

The Genome Analysis Centre

AWS Education Grant

NIH NSF Huck Institute

Penn State University Emory University

Thank you.

