

Galaxy

Daniel Blankenberg

The Galaxy Team

<http://GalaxyProject.org>

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

The Vision

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

What is Galaxy?

GUI for genomics

- ✦ for complete analyses: analyze, visualize, share, publish

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

Open source software that makes integrating your own tools and data and customizing for your own site simple

Overview

What is Galaxy?

What you can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where you can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Galaxy Analysis Workspace

The screenshot displays the Galaxy Analysis Workspace interface. The browser address bar shows <http://main.g2.bx.psu.edu/>. The main navigation bar includes 'Galaxy' and tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar contains a 'Tools' menu with categories like 'Get Data', 'Text Manipulation', 'Filter and Sort', 'Fetch Sequences', 'Get Genomic Scores', 'Statistics', 'Graph/Display Data', 'Evolution', 'Metagenomic analyses', 'EMBOSS', 'NGS TOOLBOX BETA', 'NGS: QC and manipulation', 'NGS: Mapping', 'NGS: SAM Tools', 'NGS: Indel Analysis', 'NGS: Peak Calling', 'RGENETICS', 'SNP/WGA: Data; Filters', 'SNP/WGA: QC; LD; Plots', 'SNP/WGA: Statistical Models', and 'Workflows'. The central workspace shows the 'Map with Bowtie for Illumina' tool configuration. The configuration includes: 'Use a built-in index' selected; 'mm9' selected as the reference genome; 'Paired-end' selected for library mate-pairing; '1: E18 PE.1 Reads' selected for both Forward and Reverse FASTQ files; '1000' entered for the maximum insert size; 'FR (for Illumina)' selected for mate orientation; 'Commonly used' selected for Bowtie settings; and the 'Suppress the header in the output SAM file' checkbox checked. An 'Execute' button is at the bottom of the configuration. Below the configuration is a 'What it does' section describing Bowtie as a short read aligner. The right sidebar shows a 'History' panel with a list of jobs, including '15: Variants from sample E18, consensus different, in RefSeq Genes', '14: UCSC mm9 RefSeq_Genes', '13: Variants from sample E18 where consensus base different than ref. base', '10: Variants from sample E18', '9: Generate pileup on data 8', '8: SAM-to-BAM on data 7', '7: Map with Bowtie for Illumina on data 6 and data 5', '6: E18 PE.2 Reads Groomed, Trimmed', '5: E18 PE.1 Reads Groomed, Trimmed', '4: E18 PE.2 Reads Groomed', '3: E18 PE.1 Reads Groomed', '2: E18 PE.2 Reads', and '1: E18 PE.1 Reads'. Each job entry has an eye icon, a refresh icon, and a delete icon.

Galaxy

http://main.g2.bx.psu.edu/

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools Options

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
Metagenomic analyses
EMBOSS
NGS TOOLBOX BETA
NGS: QC and manipulation
NGS: Mapping
NGS: SAM Tools
NGS: Indel Analysis
NGS: Peak Calling
RGENETICS
SNP/WGA: Data; Filters
SNP/WGA: QC; LD; Plots
SNP/WGA: Statistical Models
Workflows

Map with Bowtie for Illumina

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

Built-ins were indexed using default options

Select a reference genome:
mm9

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

Is this library mate-paired?:
Paired-end

Forward FASTQ file:
1: E18 PE.1 Reads

Must have Sanger-scaled quality values with ASCII offset 33

Reverse FASTQ file:
1: E18 PE.1 Reads

Must have Sanger-scaled quality values with ASCII offset 33

Maximum insert size for valid paired-end alignments (-X):
1000

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

Bowtie settings to use:
Commonly used

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

Execute

What it does

Bowtie is a short read aligner designed to be ultrafast and memory-efficient. It is developed by Ben Langmead and Cole Trapnell. Please cite: Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

History Options

Imported: SNP Pileup Analysis for Sample E18

15: Variants from sample E18, consensus different, in RefSeq Genes

14: UCSC mm9 RefSeq_Genes

13: Variants from sample E18 where consensus base different than ref. base

10: Variants from sample E18

9: Generate pileup on data 8

8: SAM-to-BAM on data 7

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

6: E18 PE.2 Reads Groomed, Trimmed

5: E18 PE.1 Reads Groomed, Trimmed

4: E18 PE.2 Reads Groomed

3: E18 PE.1 Reads Groomed

2: E18 PE.2 Reads

1: E18 PE.1 Reads

Filter and Sort

- Filter data on any column using simple expressions
- Sort data in ascending or descending order
- Select lines that match an

Operate on Genomic Intervals

- Intersect the intervals of two queries
- Subtract the intervals of two queries
- Merge the overlapping intervals of a query

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

Galaxy Analysis Workspace

The screenshot displays the Galaxy Analysis Workspace interface. The main window shows a workflow titled "SNP Pileup Analysis for Sample E18". The workflow steps are listed in the History panel on the right, numbered 1 through 15. The steps include: 1: E18 PE.1 Reads, 2: E18 PE.2 Reads, 3: E18 PE.1 Reads Groomed, 4: E18 PE.2 Reads Groomed, 5: E18 PE.1 Reads Groomed, Trimmed, 6: E18 PE.2 Reads Groomed, Trimmed, 7: Map with Bowtie for Illumina on data 6 and data 5, 8: SAM-to-BAM on data 7, 9: Generate pileup on data 8, 10: Variants from sample E18, 11: Variants from sample E18 where consensus base different than ref. base, 12: Variants from sample E18 where consensus base different than ref. base, 13: Variants from sample E18 where consensus base different than ref. base, 14: UCSC mm9 RefSeq_Genes, 15: Variants from sample E18, consensus different, in RefSeq Genes. The main window shows the configuration for the "Bowtie for Illumina" tool, including options for reference genome, quality values, and paired-end alignments.

Filter and Sort

- Filter data on any column using simple expressions

- Sort data in ascending or descending order

- Select lines that match an expression

Operate on Genomes

- Intersect the intervals of two queries
- Subtract the intervals of two queries
- Merge the overlapping intervals of a query

NGS: SAM Tools

- Filter SAM records by values
- Convert SAM records to BAM
- SAM-to-BAM format to BAM
- BAM-to-SAM format to SAM
- Merge BAM files together
- Generate pileup dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases

Filter pileup

Select dataset:

10: Variants from sample E18

which contains:

Pileup with six columns (simple)

See "Types of pileup datasets" below for examples

Do not consider read bases with quality lower than:

20

No variants with quality below this value will be reported

Do not report positions with coverage lower than:

3

Pileup lines with coverage lower than this value will be skipped

Only report variants?:

Yes

See "Examples 1 and 2" below for explanation

Convert coordinates to intervals?:

No

See "Output format" below for explanation

Print total number of differences?:

No

See "Example 3" below for explanation

Print quality and base string?:

Yes

See "Example 4" below for explanation

Execute

ce

aligner designed to be ultrafast and memory-efficient. It is developed by Ben Langmead, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Filter and Sort

- Filter data on any column using simple expressions

- Sort data in ascending or descending order

- Select lines that match a query

Operate on Genomes

- Intersect the intersection of two queries
- Subtract the intersection of two queries
- Merge the overlap of a query

NGS: SAM Tools

- Filter SAM on values
- Convert SAM to BAM
- SAM-to-BAM format to BAM
- BAM-to-SAM format to SAM
- Merge BAM files together
- Generate pileup dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases

Filter pileup

Select dataset:

10: Variants from sample E18

which contains:

Pileup with six columns (simple)

See "Types of pileup datasets" below for examples

Do not consider read bases with quality lower than:

20

No variants with quality below this value will be reported

Do not report positions with coverage lower than:

3

Pileup lines with coverage lower than this value will be skipped

Only report variants?:

Yes

See "Examples 1 and 2" below for explanation

Convert coordinates to intervals?:

No

See "Output format" below for explanation

Print total number of differences?:

No

See "Example 3" below for explanation

Print quality and base string?:

Yes

See "Example 4" below for explanation

Execute

aligner designed to be ultrafast and memory-efficient. It is developed by Mark Imbusch and Mark Trapnell. Please cite: Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.

History

Options



Variant Analysis for Sample E18

15: Intersect to get Variants from sample E18, consensus different, in RefSeq Genes

14: UCSC mm9 RefSeq Genes

13: Filter to get Variants from sample E18 where consensus base different than ref. base

10: Filter pileup to get Variants from sample E18

9: Generate pileup on data 8

8: SAM-to-BAM on data 7





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

6: E18 PE.2 Reads Groomed, Trimmed

5: E18 PE.1 Reads Groomed, Trimmed


User Metadata

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 ≥ 30 and quality score ≥ 20 .

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

1. Chrom	2. Start	3. End	4	5	6	7
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	



Datasources

Upload file from your computer

- ✦ FTP support for large datasets

Files directly from a sequencer

- ✦ Sample Tracking System

UCSC table browser

BioMart

interMine / modMine

EuPathDB server

EncodeDB at NHGRI

EpiGRAPH server

Tool Suites

Text Manipulation

Format Converters

Filtering and Sorting

Join, Subtract, Group

Sequence Tools

Multi-species Alignment Tools

Genomic Interval Operations

Summary Statistics

Graphing / Plotting

Regional Variation

EMBOSS

Evolution / Phylogeny

RNA-seq

ChIP-seq

GATK

Picard

RGenetics

...and more

Overview

What is Galaxy?

What you can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where you can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Data Library "Bushman"

Library Actions ▾

These are the data underlying the analyses reported in the paper "Complete Khoisan and Bantu genomes from southern Africa" by S. C. Schuster et al., published in the journal Nature, February 18, 2010. Each data set can be downloaded and/or imported into a Galaxy history. Data will be updated as the project progresses.

Name	Information	Uploaded By	Date	File Size
<input type="checkbox"/> All SNPs in personal genomes ▾	Summary table of SNPs in all individuals	greg@bx.psu.edu	2010-01-28	676.8 Mb
<input type="checkbox"/> Alu insertions in KB1 ▾		greg@bx.psu.edu	2010-02-10	14.9 Kb
<input type="checkbox"/> Alu insertions in NB1 ▾		greg@bx.psu.edu	2010-02-10	6.5 Kb
<input type="checkbox"/> KB1 microsatellites.txt ▾		greg@bx.psu.edu	2010-02-15	3.5 Mb
<input type="checkbox"/> NB1 microsatellites.txt ▾		greg@bx.psu.edu	2010-02-15	828.5 Kb
<input type="checkbox"/> amino acid differences with functional predictions ▾		greg@bx.psu.edu	2010-02-05	1.1 Mb
<input type="checkbox"/> gene copy number of CP2 and other genes in personal genomes ▾		greg@bx.psu.edu	2010-02-15	2.1 Mb
<input type="checkbox"/> indels in ABT ▾		greg@bx.psu.edu	2010-02-03	105.3 Kb
<input type="checkbox"/> indels in KB1 ▾		greg@bx.psu.edu	2010-02-03	14.2 Mb
<input type="checkbox"/> indels in MD6 ▾		greg@bx.psu.edu	2010-02-03	109.8 Kb
<input type="checkbox"/> indels in NB1 ▾		greg@bx.psu.edu	2010-02-03	515.5 Kb
<input type="checkbox"/> indels in TK1 ▾		greg@bx.psu.edu	2010-02-03	123.2 Kb
<input type="checkbox"/> novel SNPs in ABT ▾		greg@bx.psu.edu	2010-02-09	9.4 Mb
<input type="checkbox"/> novel SNPs in KB1 ▾		greg@bx.psu.edu	2010-02-09	16.9 Mb
<input type="checkbox"/> novel SNPs in MD6 ▾		greg@bx.psu.edu	2010-02-09	594.1 Kb
<input type="checkbox"/> novel SNPs in NB1 ▾		greg@bx.psu.edu	2010-02-09	4.1 Mb
<input type="checkbox"/> novel SNPs in TK1 ▾		greg@bx.psu.edu	2010-02-09	722.6 Kb
<input type="checkbox"/> sequenced exon-containing intervals ▾		greg@bx.psu.edu	2010-02-03	3.1 Mb

For selected items:

<http://usegalaxy.org/bushman>

Managing Libraries

Loading Data

- ✦ Upload a single file
- ✦ Import datasets from a Galaxy history
- ✦ Upload a directory of files
- ✦ Directly from Sequencer using Sample Tracking System

Accessing Data

- ✦ Data contents on disk are not copied
- ✦ Dataset security: public, Role-based access control (RBAC)

Annotating Library Data: Library Templates

- ✦ Build user fillable forms
- ✦ Associate at Library, Folder or Dataset level

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ **workflows**
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Galaxy Workflows

The screenshot displays the Galaxy web interface. At the top, the browser address bar shows `http://main.g2.bx.psu.edu/`. The main navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. On the left, a 'Tools' sidebar lists various categories like 'Get Data', 'Send Data', 'ENCODE Tools', 'Text Manipulation', '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', 'Metagenomic analyses', 'EMBOSS', 'NGS TOOLBOX BETA', 'NGS: QC and manipulation', 'NGS: Mapping', 'NGS: SAM Tools', 'NGS: Indel Analysis', 'NGS: Peak Calling', 'RGENETICS', 'SNP/WGA: Data; Filters', 'SNP/WGA: QC; LD; Plots', and 'SNP/WGA: Statistical Models'. The central panel shows a table of genomic data with a warning message: 'This dataset is large and only the first megabyte is shown below.' The table columns include chromosome (chr10), coordinates, and various metrics. On the right, a 'History' panel shows a list of workflow steps, including 'Generate pileup on data 8', 'SAM-to-BAM on data Z', and 'Map with Bowtie for Illumina on data 6 and data 5'. A context menu is open over the history list, showing options like 'Create New', 'Clone', 'Share or Publish', 'Extract Workflow', 'Dataset Security', 'Show Deleted Datasets', 'Show Hidden Datasets', 'Show structure', and 'Delete'.

Galaxy Workflows

The screenshot displays the Galaxy workflow editor interface. On the left, a sidebar lists various tool categories such as 'Get Data', 'Text Manipulation', 'FASTA manipulation', and 'NGS TOOLBOX BETA'. The main workspace is divided into two columns: 'Tool' and 'History items created'. The 'Tool' column contains several tool entries, each with a checkbox to include it in the workflow. The 'History items created' column shows the resulting workflow steps, numbered 1 through 9. A right-hand panel shows a 'History Lists' menu with options like 'Extract Workflow' and 'Delete'.

Tool	History items created
Upload File <i>This tool cannot be used in workflows</i>	1: E18 PE.1 Reads <input checked="" type="checkbox"/> Treat as input dataset
Upload File <i>This tool cannot be used in workflows</i>	2: E18 PE.2 Reads <input checked="" type="checkbox"/> Treat as input dataset
FASTQ Groomer <input checked="" type="checkbox"/> Include "FASTQ Groomer" in workflow	3: E18 PE.1 Reads Groomed
FASTQ Groomer <input checked="" type="checkbox"/> Include "FASTQ Groomer" in workflow	4: E18 PE.2 Reads Groomed
FASTQ Trimmer <input checked="" type="checkbox"/> Include "FASTQ Trimmer" in workflow	5: E18 PE.1 Reads Groomed, Trimmed
FASTQ Trimmer <input checked="" type="checkbox"/> Include "FASTQ Trimmer" in workflow	6: E18 PE.2 Reads Groomed, Trimmed
Map with Bowtie for Illumina <input checked="" type="checkbox"/> Include "Map with Bowtie for Illumina" in workflow	7: Map with Bowtie for Illumina on data 6 and data 5
SAM-to-BAM <input checked="" type="checkbox"/> Include "SAM-to-BAM" in workflow	8: SAM-to-BAM on data 7
Generate pileup <input checked="" type="checkbox"/> Include "Generate pileup" in workflow	9: Generate pileup on data 8

Galaxy Workflows

The screenshot displays the Galaxy workflow editor interface. The browser address bar shows the URL: <http://main.g2.bx.psu.edu/workflow/editor?id=a6d94f12f42c1af8>. The page title is "Galaxy" and the navigation menu includes "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User".

The main workspace is titled "Workflow Canvas | SNP variant detection from paired-end reads". It features a grid background with several workflow steps connected by lines:

- Input dataset** (output) connects to **FASTQ Groomer** (File to groom, output_file: fastqsanger, fastqcssanger, fastqsolexa, fastqillumina).
- FASTQ Groomer** connects to **FASTQ Trimmer** (FASTQ File, output_file).
- FASTQ Trimmer** connects to **Map with Bowtie for Illumina** (Forward FASTQ file, Reverse FASTQ file, output (sam)).
- Map with Bowtie for Illumina** connects to **SAM-to-BAM** (SAM File to Convert, output1 (bam)).
- SAM-to-BAM** connects to **Generate pileup** (Select the BAM file to generate the pileup file for, output1 (tabular)).
- Generate pileup** connects to **Filter pileup** (Select dataset, out_file1 (tabular, interval)).

A sidebar on the left lists various tools, including "Get Data", "Send Data", "ENCODING", "Lift-Over", "Text Manipulation", "Conversion", "FASTA", "Filter", "Join", "Subtract", "Extract", "Fetch", "Fetch A", "Get Gen", "Operate", "Statistic", "Graph/Region", "Multiple", "Multivariate", "Evolution", "Metagenomics", "EMBOSS", "NGS TO", "NGS: QC", "NGS: M", "NGS: SA", "NGS: In", "NGS: Pe", "RGENET", "SNP/WC", "SNP/WC", "SNP/WC", and "Workflow".

At the bottom of the canvas, there is a checkbox labeled "Include 'Generate pileup' in workflow" which is checked.

Galaxy Workflows

The screenshot displays the Galaxy workflow editor interface. The main window shows a workflow canvas titled "Workflow Canvas | SNP variant detection from paired-end reads". The workflow consists of several interconnected steps:

- Input dataset** (output) feeds into a **FASTQ Groomer** step (File to groom).
- The **FASTQ Groomer** step feeds into a **FASTQ Trimmer** step (FASTQ File).
- The **FASTQ Trimmer** step feeds into a **Map with Bowtie for Illumina** step (Forward FASTQ file and Reverse FASTQ file).
- The **Map with Bowtie for Illumina** step feeds into a **SAM-to-BAM** step (SAM File to Convert).
- The **SAM-to-BAM** step feeds into a **Generate pileup** step (Select generate pileup).

On the right side, a tool configuration panel is visible for the **Tool: SAM-to-BAM** step. It includes the following options:

- Choose the source for the reference list:**
- SAM File to Convert:** Data input 'input1' (sam)
- Edit Step Actions:**
- Edit Step Attributes:**

At the bottom of the interface, there is a checkbox labeled "Include 'Generate pileup' in workflow" which is currently checked.

Galaxy Workflows

The image displays the Galaxy Workflows interface. In the background, a workflow canvas titled "Workflow Canvas | SNP variant dete" is visible, showing a sequence of steps: "Input dataset" (output), "FASTQ Groomer" (File to groom, output_file (fastqsanger, fastq...)), "FASTQ Trimmer" (output_file), "FASTQ File" (output_file), and "FASTQ Groomer" (File to groom, output_file (fastqsanger, fastq...)).

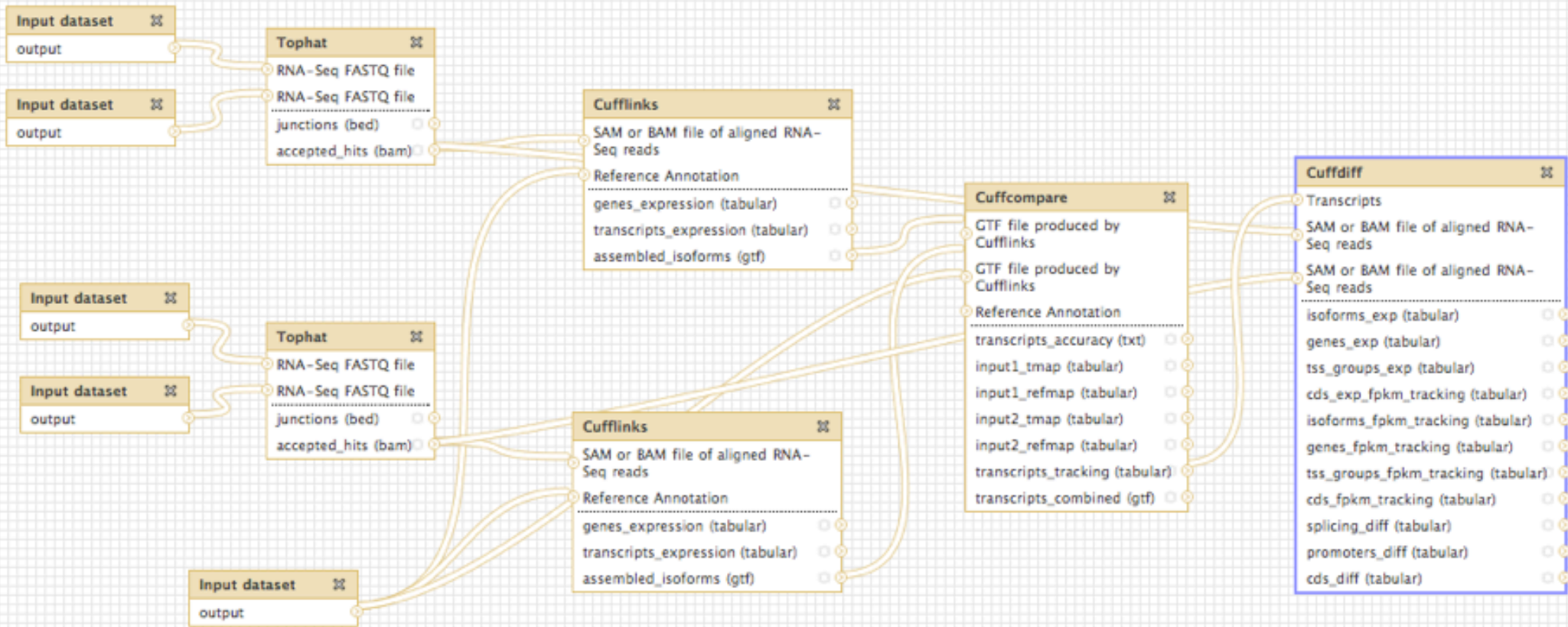
Overlaid on the canvas is the "Edit Workflow Attributes" dialog box. It contains the following information:

- Name:** SNP identification within annotated genes from NGS PE Data
- Tags:** snp x ngs x pileup x bowtie x
- Annotation / Notes:** Identify variants in annotated genes from NGS paired-end data.
- Apply tags to make it easy to search for and find items with the same tag.**
- Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.**

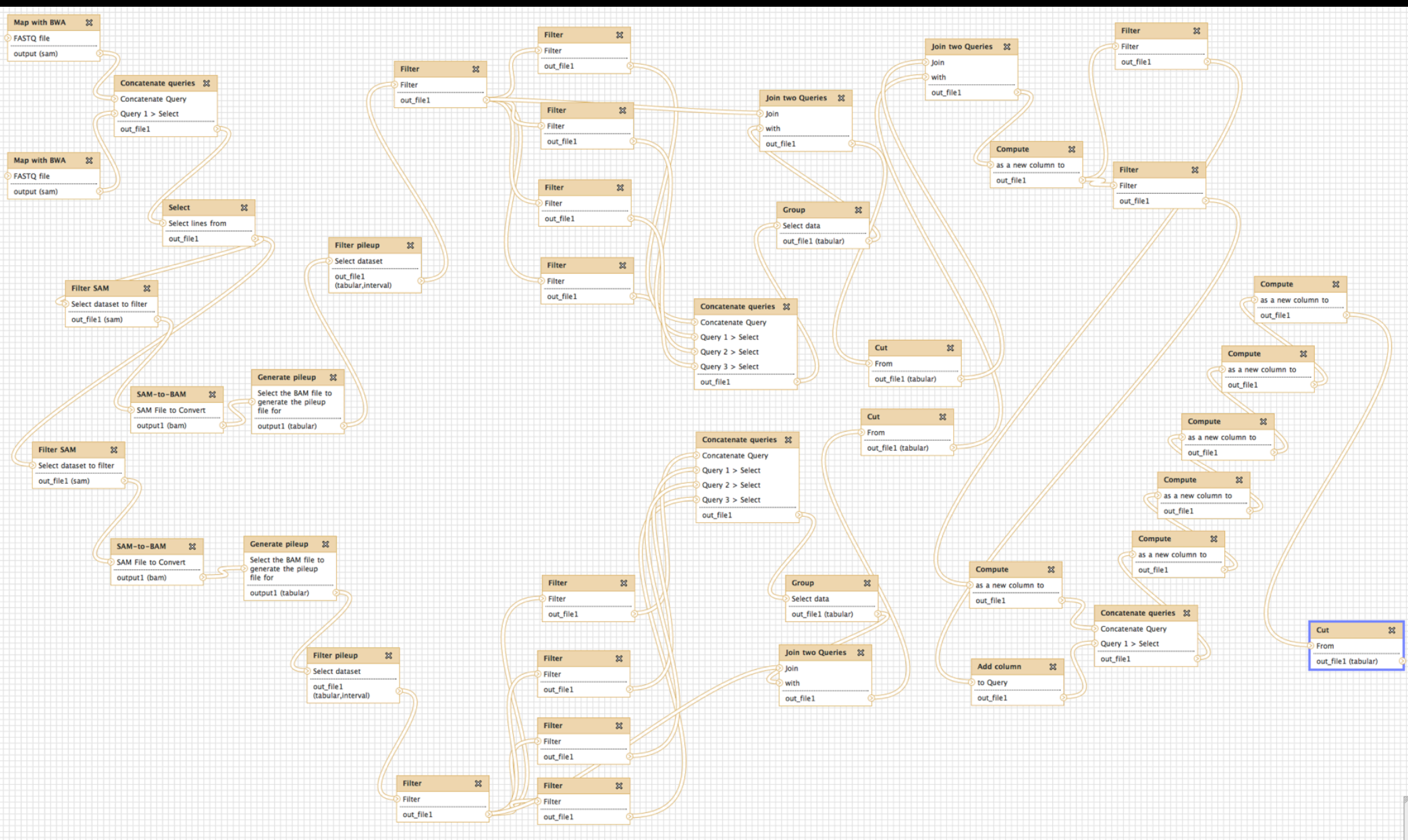
To the right of the dialog is the "Tool: SAM-to-BAM" configuration panel:

- Choose the source for the reference list:** Locally cached
- SAM File to Convert:** Data input 'input1' (sam)
- Edit Step Actions:** Assign Columns, output1, Create
- Edit Step Attributes:** Annotation / Notes: Convert Bowtie SAM output to BAM format so that pileup can be run.

At the bottom of the interface, a checkbox is checked: "Include 'Generate pileup' in workflow".



Example: Workflow for differential expression analysis of RNA-seq using Tophat/Cufflinks tools



Example: Diagnosing low-frequency heterosplasmic sites in two tissues from the same individual

Overview

What is Galaxy?

What you can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ **visualization**
- ✦ sharing
- ✦ Pages

Where you can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Visualize

Send data results to external genome browsers

Trackster: Galaxy's genome browser

External Genome Browsers

UCSC

Ensembl

GBrowse






IGV

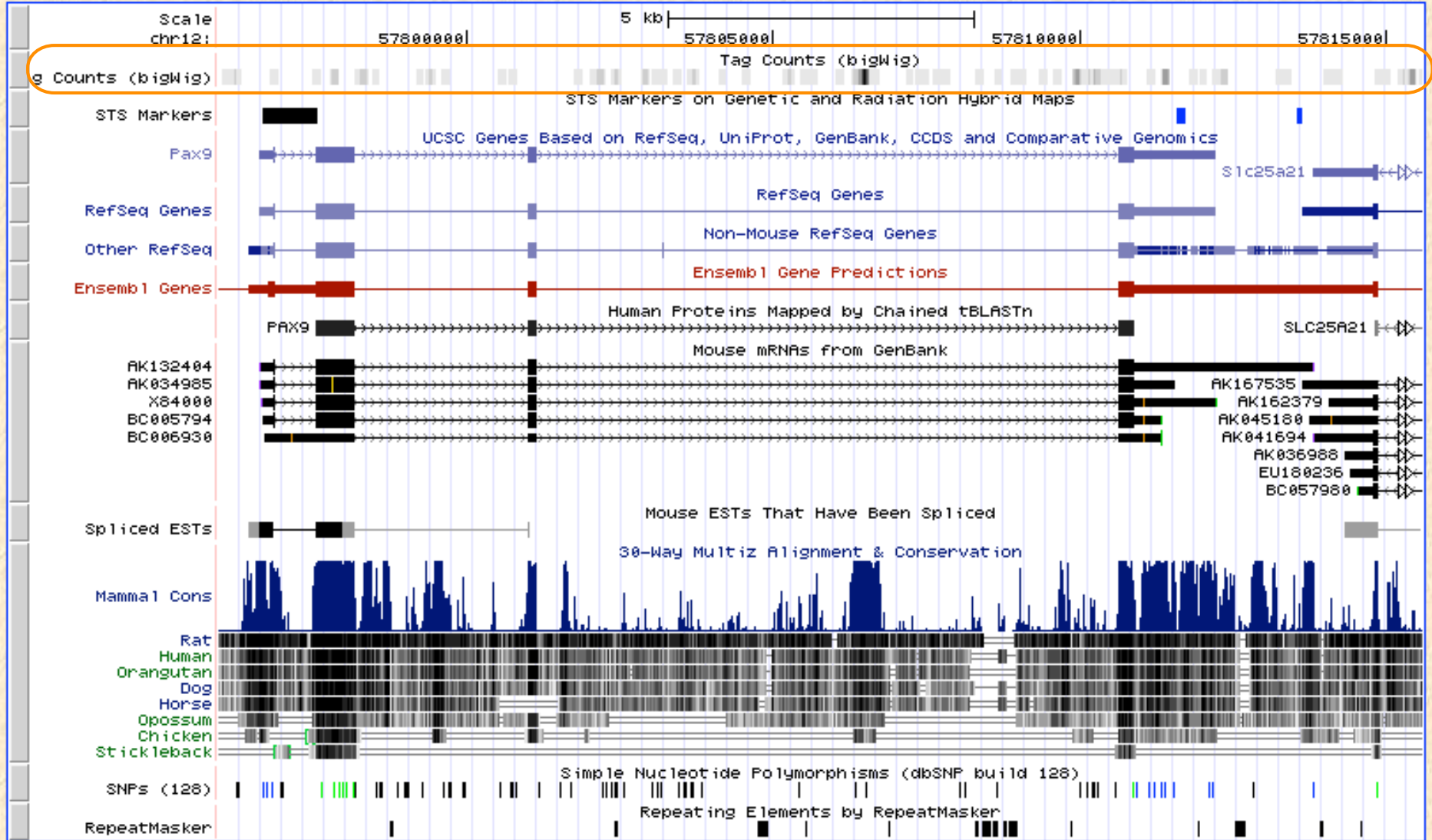
UCSC Genome Browser on Mouse July 2007 (NCBI37)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out

position/search chr12:57,795,963-57,815,592 gene jump clear size 17,000 bp. compare

chr12 (qC1) 12qA1.1 qA2 12qA3 qB1 12qB3 12qC1 qC2 12qC3 qD1 qD2 12qD3 12qE 12qF1 qF2

14: Tag Counts (bigWig)   
2.4 Gb, format: bigwig, database: mm9
Info:  
[display at UCSC main](#)
Binary UCSC BigWig file



Integrative Genomics Viewer (IGV)

1: Sample data

1.2 Gb
format: bam, database: mm9
Info: uploaded bam file



display at UCSC [main](#) [test](#)
display at Ensembl [Current](#)
display with IGV [web](#) [local](#)

Binary bam alignments file



The application "IGV 1.5" from "www.broadinstitute.org" is requesting access to your computer.

The digital signature could not be verified.

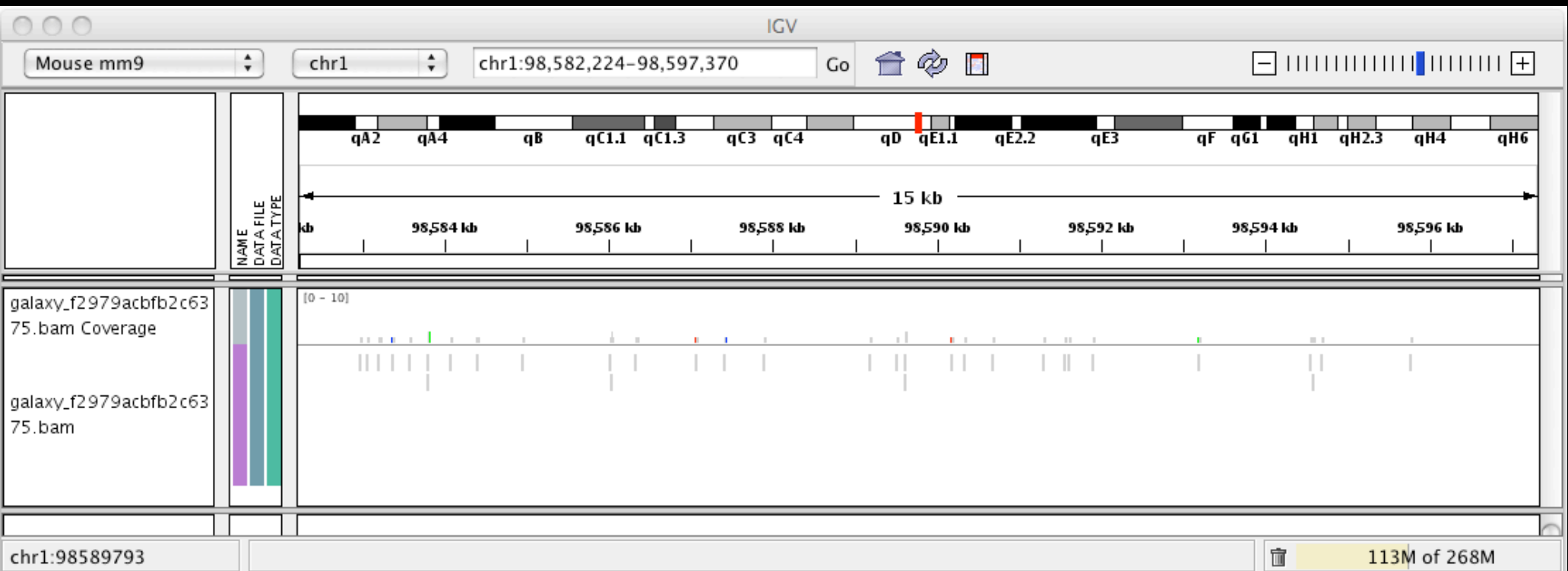
Allow all applications from "www.broadinstitute.org" with this signature



Show Details...

Deny

Allow



Galaxy

- ✦ tool integration framework
- ✦ heavy focus on usability
- ✦ sharing, publication framework

Genome Browser

- ✦ physical depiction of data
- ✦ visually identify correlations
- ✦ find interesting regions, features

Galaxy

- ✦ tool integration framework
- ✦ heavy focus on usability
- ✦ sharing, publication framework

Genome Browser

- ✦ physical depiction of data
- ✦ visually identify correlations
- ✦ find interesting regions, features



Trackster

The diagram consists of three light blue rounded rectangular boxes on a black background. On the left, there are two boxes stacked vertically. The top box is titled 'Galaxy' and contains three bullet points: 'tool integration framework', 'heavy focus on usability', and 'sharing, publication framework'. The bottom box is titled 'Genome Browser' and contains three bullet points: 'physical depiction of data', 'visually identify correlations', and 'find interesting regions, features'. On the right, there is a larger box titled 'Trackster'. Two orange curved arrows point from the right side of the 'Galaxy' box and the right side of the 'Genome Browser' box towards the 'Trackster' box, indicating a relationship or integration between these tools and Trackster.

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 1,290 - 4,168,475

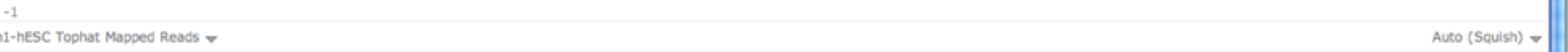
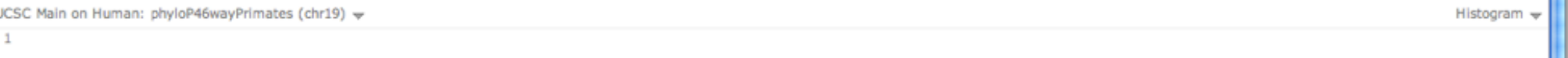
0 1,000,000 2,000,000 3,000,000 4,000,000



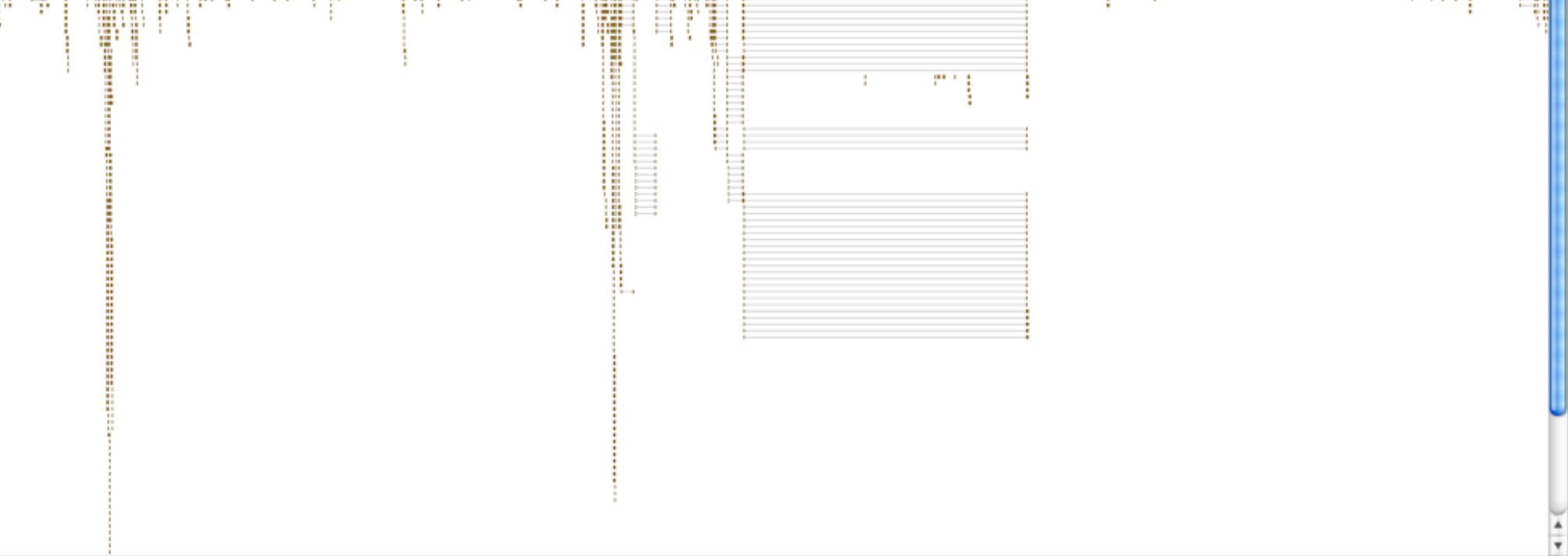
0 1,000,000 2,000,000 3,000,000 4,000,000

Published Visualizations | jeremy | GCC2011-1: Viewing and chr19 625,719 - 682,581

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



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 g c c e g g g e c T C A C C G G C A G G C G C G G G A 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 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)

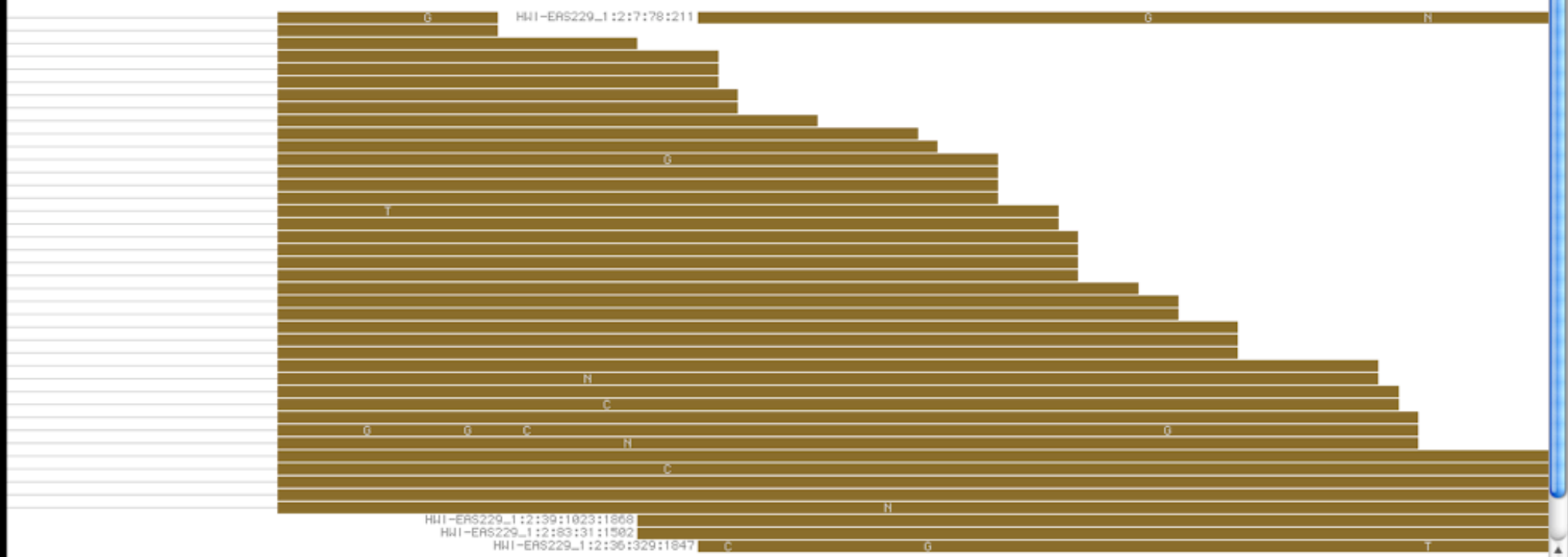
UCSC Main on Human: all_est (chr19) Dense



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

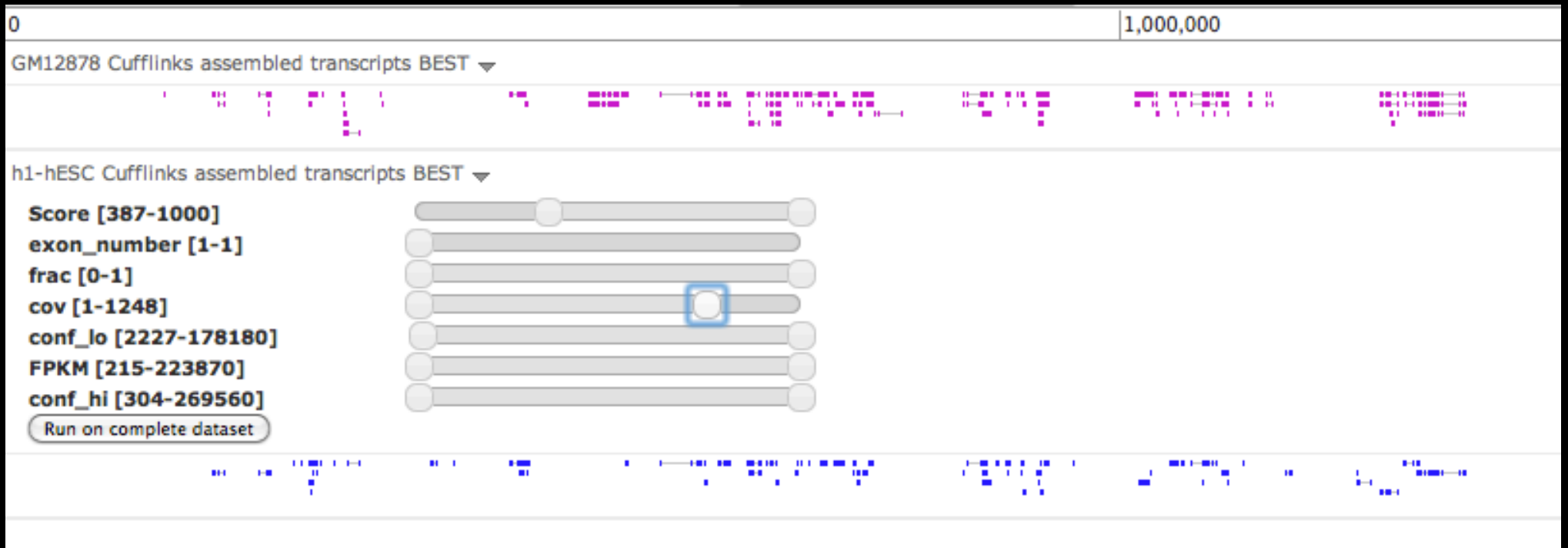
But really, why *another* genome browser

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



Integrating Tools and Visualization

Galaxy Analyze Data Workflow Shared Data **Visualization** Admin Help User

GCC3: Running Tools (hg19) chr19 1,523,098 - 1,545,232 1,530,000 1,540,000

UCSC Main on Human: knownGene

h1-hESC Tophat mapped reads

h1-hESC assembled transcripts - region=[all], parameters=[150000, 0.5, 0.05, No]

Cufflinks

Max Intron Length: 150000

Min Isoform Fraction: 0.5

Pre MRNA Fraction: 0.05

Perform quartile normalization: No

Run on complete dataset Run on visible region

FF .138.1 CUFF .139.1 CUFF .140.1 CUFF .141.1 CUFF .142.1

h1-hESC assembled transcripts - region=[all], parameters=[150000, 0.5, 0.05, No] ▼

Cufflinks

Max Intron Length

150000

Min Isoform Fraction

0.05

Pre MRNA Fraction

0.05

Perform quartile normalization

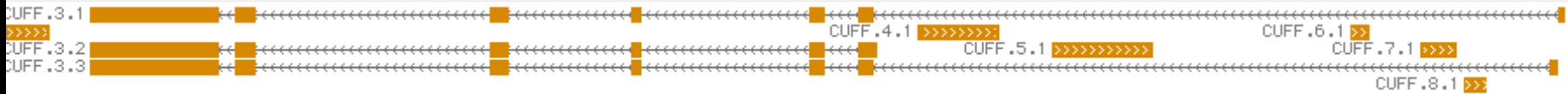
No ▼

Run on complete dataset

Run on visible region



→ Cufflinks - region=[chr19:1523098-1545232], parameters=[150000, 0.05, 0.05, No] ▼



Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ **sharing**
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Sharing and Publishing

Sharing and Publishing History 'Variant Analysis for Sample E18'

Making History Accessible via Link and Publishing It

This history is currently restricted so that only you and the users listed below can access it. You can:

Make History Accessible via Link

Generates a web link that you can share with other people so that they can view and import the history.

Make History Accessible and Publish

Makes the history accessible via link (see above) and publishes the history to Galaxy's [Published Histories](#) section, where it is publicly listed and searchable.

Sharing History with Specific Users

You have not shared this history with any users.

Share with a user

[Back to Histories List](#)

Sharing and Publishing

Sharing and Publishing History 'Variant Analysis for Sample E18'

Making History Accessible via Link and Publishing It

This history accessible via link and published.

Anyone can view and import this history by visiting the following URL:

<http://main.g2.bx.psu.edu/u/jgoecks/h/variant-analysis-for-sample-e18> 

This history is publicly listed and searchable in Galaxy's Published Histories section.

You can:

Unpublish History

Removes history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

Disable Access to History via Link and Unpublish

Disables history's link so that it is not accessible and removes history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

Sharing History with Specific Users

You have not shared this history with any users.

Share with a user

[Back to Histories List](#)

Galaxy | Published History | Variant Analysis for Sample E18

http://main.g2.bx.psu.edu/u/jgoecks/h/variant-analysis-for-sample-e18

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Published Histories | jgoecks | Variant Analysis for Sample E18


Galaxy History ' Variant Analysis for Sample E18'

[+ Import history](#)

Annotation: Perform a pileup analysis with default parameters to identify variants in sample E18.

Dataset	Annotation
1: E18 PE.1 Reads	Forward reads from sample E18.
2: E18 PE.2 Reads	Reverse reads from sample E18.
3: E18 PE.1 Reads Groomed	Groom reads to convert quality scores from Solexa 1.0 to Solexa 1.3
4: E18 PE.2 Reads Groomed	Groom reads to convert quality scores from Solexa 1.0 to Solexa 1.3
5: E18 PE.1 Reads Groomed, Trimmed	Trim reads from 3' end to remove low-quality nts.
6: E18 PE.2 Reads Groomed, Trimmed	Trim reads from 3' to remove low-quality nts.
7: Map with Bowtie for Illumina on data 6 and data 5	Map paired-end reads with default parameters.
8: SAM-to-BAM on data 7	Need to convert Bowtie SAM to BAM so that pileup analysis can be performed.
9: Generate pileup on data 8	Pileup analysis with default parameters
10: Filter pileup to get Variants from sample E18	Find variants with coverage ≥ 30 .
13: Filter to get Variants from sample E18 where consensus base different than ref. base	Filter pileup to find variants where the consensus base is different than the reference base.
14: UCSC mm9 RefSeq Genes	UCSC mm9 RefSeq genes.
15: Intersect to get Variants from sample E18, consensus different, in RefSeq Genes	Variants with consensus different that occur in RefSeq genes.

About this History

Author  jgoecks

Related Histories
[All published histories](#)
[Published histories by jgoecks](#)

Rating
 Community (1 rating, 4.0 average) ★ ★ ★ ★ ☆
 Yours ★ ★ ★ ★ ☆

Tags
 Community: snp pileup bowtie demo
sample
 Yours: snp pileup bowtie
demo sample:e18 +

Galaxy | Published Workflow | SNP variant detection from paired-end reads

http://main.g2.bx.psu.edu/u/jgoecks/w/snp-variant-detection-from-paired-end-reads

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Published Workflows | jgoecks | SNP variant detection from paired-end reads

Step 6: FASTQ Trimmer

Trim reads to remove low-quality bases.

FASTQ File
Output dataset 'output_file' from step 4

Define Base Offsets as
Absolute Values

Offset from 5' end
0

Offset from 3' end
9

Keep reads with zero length
False

Step 7: Map with Bowtie for Illumina

Map reads using default parameter values.

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

Select a reference genome
/galaxy/data/apiMe13/bowtie_index/apiMe13

Is this library mate-paired?
Paired-end

Forward FASTQ file
Output dataset 'output_file' from step 6

Reverse FASTQ file
Output dataset 'output_file' from step 5

Maximum insert size for valid paired-end alignments (-X)
1000

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

Bowtie settings to use
Commonly used


Suppress the header in the output SAM file
True

Step 8: SAM-to-BAM

Convert Bowtie SAM output to BAM format so that pileup can be run.

Choose the source for the reference list
Locally cached

About this Workflow

Author
jgoecks 

Related Workflows
[All published workflows](#)
[Published workflows by jgoecks](#)

Rating
Community (0 ratings, 0.0 average) ★★★★★
Yours ★★★★★

Tags
Community:
snp bowtie
Yours:
snp bowtie

Published Histories

[Advanced Search](#)

Name	Annotation	Owner	Community Rating ↑	Community Tags	Last Updated
Galaxy vs MEGAN	Comparison of Galaxy vs. MEGAN pipeline.	aun1	★★★★★	metagenomics megan galaxy	Mar 19, 2010
metagenomic analysis		aun1	★★★★★	metagenomics galaxy	Mar 19, 2010
SM_1186088	Datasets correspond to our paper published in Science by Peleg et al. entitled : Altered histone acetylation is associated with age-dependent memory impairment. Experiment layout: This history contains 4 datasets in the form of BED files of uniquely mapped reads produced after chip-seq for histone modifications H4K12ac and H3K9ac in mouse hippocampus of 3 months (young) and 16 months (old) mice after fear conditioning. For detailed information please refer to supplementary materials and methods of the respective work by peleg et al.	fischerlab	★★★★★		Apr 19, 2010
Variant Analysis for Sample E18	Perform a pileup analysis with default parameters to identify variants in sample E18.	jgoecks	★★★★★	snp pileup bowtie demo sample	2 minutes ago
get longest exon		henri	★★★★★	chr22 longest marc exon human workshop	Sep 02, 2010
FASTA to Tabular Test		JJ	★★★★★		Aug 26, 2010
EKLF		yzc109	★★★★★		Aug 24, 2010

Open "http://main.g2.bx.psu.edu/history/list_published?sort=rating&f-tags=All" in a new tab

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ **Pages**

Where **you** can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Galaxy Pages

A web-based, interactive medium for presenting all aspects of an analysis: data, methods, and results

Galaxy Pages

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/jgoecks/p/variant-analysis-for-sample-e18>. The page title is "Galaxy | Published Page | Variant Analysis for sample E18". The main content area features a header "Variant Analysis of Embryonic Mouse Brain Tissue" by Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team. The "Results" section describes an NGS re-sequencing experiment on mm9 brain tissue, identifying 27,742 possible variants, with 5,625 having a consensus base different from the reference and 2,796 occurring in known RefSeq Genes. A green highlighted box contains a link to a "Galaxy Dataset | Intersect to get Variants from sample E18, consensus different, in RefSeq Genes". The "Method" section details the bioinformatics pipeline, including read grooming, mapping with Bowtie, and pileup analysis with SAMtools. A blue highlighted box contains a link to the "Galaxy History | Variant Analysis for Sample E18". Below this, a workflow for performing the analysis is shown, with an orange highlighted box containing a link to a "Galaxy Workflow | Variant Identification within annotated genes from NGS PE Data". The "References" section lists two papers: [1] Han, X. et al. (2009) and [2] Langmead, B. et al. (2009). On the right side, the "About this Page" sidebar shows the author's profile (jgoecks), related pages, a rating of 0 stars, and tags.

Galaxy | Published Page | Variant Analysis for sample E18

Published Pages | jgoecks | Variant Analysis for sample E18

Variant Analysis of Embryonic Mouse Brain Tissue

Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team

Results

To demonstrate how Galaxy can support accessible, reproducible, and transparent NGS re-sequencing studies, we performed a simple variant analysis experiment. This experiment identifies variants from a set of 4,536,964 RNA-seq reads obtained from sequencing a sample of mm9 brain tissue from day 18 of embryonic development.

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2,796 occur in known RefSeq Genes. These potential variants are:

[Galaxy Dataset | Intersect to get Variants from sample E18, consensus different, in RefSeq Genes](#)
Variants with consensus different that occur in RefSeq genes.

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper Bowtie [2]. A pileup analysis using SAMtools [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

[Galaxy History | Variant Analysis for Sample E18](#)
Perform a pileup analysis with default parameters to identify variants in sample E18.

Here is a workflow for performing this analysis:

[Galaxy Workflow | Variant Identification within annotated genes from NGS PE Data](#)
Identify variants in annotated genes from NGS paired-end data.

References

[1] Han, X. et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proceedings of the National Academy of Sciences* **106**, 12741-12746 (2009).

[2] Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

[3] Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

About this Page

Author: jgoecks

Related Pages: [All published pages](#), [Published pages by jgoecks](#)

Rating: Community (0 ratings, 0.0 average), Yours

Tags: Community: none, Yours

Galaxy Pages

Galaxy | Published Page | Variant Analysis for sample E18

http://main.g2.bx.psu.edu/u/jgoecks/p/variant-analysis-for-sample-e18

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Published Pages | jgoecks | Variant Analysis for sample E18

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2796 occur in known RefSeq Genes. These potential variants are:

[Galaxy Dataset | Intersect to get Variants from sample E18, consensus different, in RefSeq Genes](#)
Variants with consensus different that occur in RefSeq genes.

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper Bowtie [2]. A pileup analysis using SAMtools [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

Galaxy History | Variant Analysis for Sample E18
Perform a pileup analysis with default parameters to identify variants in sample E18.

8: SAM-to-BAM on data 7	Need to convert Bowtie SAM to BAM so that pileup analysis can be performed.
9: Generate pileup on data 8	Pileup analysis with default parameters
10: Filter pileup to get Variants from sample E18	Find variants with coverage ≥ 30 .
13: Filter to get Variants from sample E18 where consensus base different than ref. base	Filter pileup to find variants where the consensus base is different than the reference base.
14: UCSC mm9 RefSeq Genes	UCSC mm9 RefSeq genes.
15: Intersect to get Variants from sample E18, consensus different, in RefSeq Genes	Variants with consensus different that occur in RefSeq genes.

Here is a workflow for performing this analysis:

[Galaxy Workflow | Variant identification within annotated genes from NGS PE Data](#)
Identify variants in annotated genes from NGS paired-end data.

References

About this Page

Author
jgoecks

Related Pages
[All published pages](#)
[Published pages by jgoecks](#)

Rating
Community (0 ratings, 0.0 average) ★★★★★
Yours ★★★★★

Tags
Community: none
Yours:

Galaxy Pages

Galaxy | Published Page | Variant Analysis for sample E18

http://main.g2.bx.psu.edu/u/jgoecks/p/variant-analysis-for-sample-e18

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Published Pages | jgoecks | Variant Analysis for sample E18

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2796 occur in known RefSeq Genes. These potential variants are:

[Galaxy Dataset | Intersect to get Variants from sample E18, consensus different, in RefSeq Genes](#)
Variants with consensus different that occur in RefSeq genes.

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper Bowtie [2]. A pileup analysis using SAMtools [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

Galaxy History | Variant Analysis for Sample E18
Perform a pileup analysis with default parameters to identify variants in sample E18.

8: SAM-to-BAM on data 7	Need to convert Bowtie SAM to BAM so that pileup analysis can be performed.
9: Generate pileup on data 8	Pileup analysis with default parameters
10: Filter pileup to get Variants from sample E18	Find variants with coverage >= 30.
13: Filter to get Variants from sample E18 where consensus base different than ref. base	Filter pileup to find variants where the consensus base is different than the reference base.
14: UCSC mm9 RefSeq Genes	UCSC mm9 RefSeq genes.
15: Intersect to get Variants from sample E18, consensus different, in RefSeq Genes	Variants with consensus different that occur in RefSeq genes.

Here is a workflow for performing this analysis:

[Galaxy Workflow | Variant identification within annotated genes from NGS PE Data](#)
Identify variants in annotated genes from NGS paired-end data.

References

Open "http://main.g2.bx.psu.edu/history/imp?id=e0b8bd5d661b10c2" in a new tab

About this Page

Author: jgoecks

Related Pages: [All published pages](#), [Published pages by jgoecks](#)

Rating: Community (0 ratings, 0.0 average) ★★★★★, Yours ★★★★★

Tags: Community: none, Yours:

Galaxy Pages

The screenshot shows the Galaxy web interface. The browser address bar displays `http://main.g2.bx.psu.edu/`. The main navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar lists various tool categories such as 'Get Data', 'Text Manipulation', 'Filter and Sort', and 'NGS: QC and manipulation'. The central workspace displays the 'Filter pileup' tool configuration page. The tool is set to '9: Generate pileup on data 8' and 'Pileup with ten columns (with consensus)'. It includes input fields for 'Do not consider read bases with quality lower than:' (20) and 'Do not report positions with coverage lower than:' (30). The 'Only report variants?' and 'Convert coordinates to intervals?' options are set to 'Yes'. The 'Print total number of differences?' and 'Print quality and base string?' options are also set to 'Yes'. An 'Execute' button is visible at the bottom of the tool configuration. The right sidebar shows a 'History' panel with a list of recent jobs, including '15: Variants from sample E18, consensus different, in RefSeq Genes', '14: UCSC mm9 RefSeq Genes', '13: Variants from sample E18 where consensus base different than ref. base', '10: Variants from sample E18', '9: Generate pileup on data 8', '8: SAM-to-BAM on data Z', '7: Map with Bowtie for Illumina on data 6 and data 5', and '6: E18 PE.2 Reads'. A table of genomic coordinates is visible in the history panel for job 10.

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	X	173
chr10	14465083	14465084	G	X	144
chr10	14465084	14465085	T	T	117

Open `http://main.g2.bx.psu.edu/tool_runner/rerun?id=1703758` in a new tab

Galaxy Pages

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/jgoecks/p/variant-analysis-for-sample-e18>. The page title is "Variant Analysis of Embryonic Mouse Brain Tissue" by Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team. The page is divided into sections: "Results", "Method", and "References".

Results

To demonstrate how Galaxy can support accessible, reproducible, and transparent NGS re-sequencing studies, we performed a simple variant analysis experiment. This experiment identifies variants from a set of 4,536,964 RNA-seq reads obtained from sequencing a sample of mm9 brain tissue from day 18 of embryonic development.

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2796 occur in known RefSeq Genes. These potential variants are:

[Galaxy Dataset | Intersect to get Variants from sample E18, consensus different, in RefSeq Genes](#)
Variants with consensus different that occur in RefSeq genes.

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper Bowtie [2]. A pileup analysis using SAMtools [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

[Galaxy History | Variant Analysis for Sample E18](#)
Perform a pileup analysis with default parameters to identify variants in sample E18.

Here is a workflow for performing this analysis:

[Galaxy Workflow | Variant Identification within annotated genes from NGS PE Data](#)
Identify variants in annotated genes from NGS paired-end data.

References

[1] Han, X. et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proceedings of the National Academy of Sciences* 106, 12741-12746 (2009).

[2] Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

[3] Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).

About this Page

Author
jgoecks

Related Pages
[All published pages](#)
[Published pages by jgoecks](#)

Rating
Community (0 ratings, 0.0 average) ★★★★★
Yours ★★★★★

Tags
Community: none
Yours:

Creating a Page

The screenshot shows a web browser window with the URL `http://main.g2.bx.psu.edu/page/edit_content?id=d2523e005e1ec427`. The browser title is "Galaxy". The page header includes the Galaxy logo and navigation links: "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User". The page editor interface shows the title "Variant Analysis of Embryonic Mouse Brain Tissue" and the authors "Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team". The page content is divided into sections: "Results", "Method", and "References".

Page Editor | Title : Variant Analysis for sample E18 Save Close

B I \times^2 \times_2 Paragraph type ▾ Insert Link to Galaxy Object ▾ Embed Galaxy Object ▾

Variant Analysis of Embryonic Mouse Brain Tissue

Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team

Results

To demonstrate how Galaxy can support accessible, reproducible, and transparent NGS re-sequencing studies, we performed a simple variant analysis experiment. This experiment identifies variants from a set of 4,536,964 RNA-seq reads obtained from sequencing a sample of mm9 brain tissue from day 18 of embryonic development.

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2796 occur in known RefSeq Genes. These potential variants are:

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper [Bowtie](#) [2]. A pileup analysis using [SAMtools](#) [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

Here is a workflow for performing this analysis:

References

[1] Han, X. et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proceedings of the National Academy of Sciences* **106**, 12741-12746 (2009).

[2] Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

Creating a Page

Galaxy

Page Editor | Title : Variant Analysis for sample E18

Variant Analysis of Embryonic Development

Jeremy Goecks, Anton Nekrutenko, James Taylor, et al.

Results

To demonstrate how Galaxy can support accessible, identifies variants from a set of 4,536,964 RNA-seq

The initial analysis produced support for 27,742 positions from the reference base and (b) read coverage at the

Method

In the first step of this analysis, the reads were grouped to exclude base pairs with low quality scores; see [1] for [Bowtie](#) [2]. A pileup analysis using SAMtools [3] was

variant analysis experiment. This experiment bryonic development.

base--as determined by the MAQ model--differs genes. These potential variants are:

it, the reads were trimmed from 36bp to 27bp to reads were mapped using the short-read mapper complete analysis is contained in this history:

References

[1] Han, X. et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proceedings of the National Academy of Sciences* 106, 12741-12746 (2009).

[2] Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

Name	Tags	Last Updated ↑
<input checked="" type="checkbox"/> Variant Analysis for Sample E18	5 Tags	15 minutes ago
<input type="checkbox"/> Pileup analysis, sample E18	4 Tags	2 days ago
<input type="checkbox"/> Unnamed history	0 Tags	Sep 07, 2010
<input type="checkbox"/> Unnamed history	0 Tags	Dec 17, 2009
<input type="checkbox"/> imported: Hsitory with ~100 items	5 Tags	Dec 10, 2009
<input type="checkbox"/> imported: Galaxy vs MEGAN	0 Tags	Dec 04, 2009
<input type="checkbox"/> imported: Galaxy vs MEGAN	2 Tags	Oct 06, 2009
<input type="checkbox"/> imported: Galaxy vs MEGAN	0 Tags	Oct 06, 2009
<input type="checkbox"/> imported: metagenomic analysis	0 Tags	Sep 30, 2009
<input type="checkbox"/> imported: Galaxy vs MEGAN	0 Tags	Sep 30, 2009

Page: 1 2 | [Show all histories on one page](#)

For 1 selected histories:

Make the selected histories accessible so that they can viewed by everyone.

[Embed](#) [Cancel](#)

Creating a Page

The screenshot shows a web browser window with the URL `http://main.g2.bx.psu.edu/page/edit_content?id=d2523e005e1ec427`. The browser title is "Galaxy". The page header includes navigation links: "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User". The main content area is titled "Page Editor | Title : Variant Analysis for sample E18" and includes "Save" and "Close" buttons. Below the header is a rich text editor toolbar with options for "Paragraph type", "Insert Link to Galaxy Object", and "Embed Galaxy Object".

Variant Analysis of Embryonic Mouse Brain Tissue

Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team

Results

To demonstrate how Galaxy can support accessible, reproducible, and transparent NGS re-sequencing studies, we performed a simple variant analysis experiment. This experiment identifies variants from a set of 4,536,964 RNA-seq reads obtained from sequencing a sample of mm9 brain tissue from day 18 of embryonic development.

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2796 occur in known RefSeq Genes. These potential variants are:

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper [Bowtie](#) [2]. A pileup analysis using [SAMtools](#) [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

Embedded Galaxy History 'Variant Analysis for Sample E18'

[Do not edit this block; Galaxy will fill it in with the annotated history when it is displayed.]

Here is a workflow for performing this analysis:

References

[1] Han, X. et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proceedings of the National Academy of Sciences* **106**, 12741-12746 (2009).

Open # on this page in a new tab

Creating a Page

The screenshot shows the Galaxy web interface in a browser window. The address bar displays the URL `http://main.g2.bx.psu.edu/page/edit_content?id=d2523e005e1ec427`. The Galaxy logo is in the top left, and navigation tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User' are in the top right. The page title is 'Page Editor | Title : Variant Analysis for sample E18'. A rich text editor toolbar is visible with options for bold, italic, text color, background color, bulleted list, numbered list, link, unlink, undo, redo, insert link to Galaxy object, and embed Galaxy object. The main content area contains the following text:

To demonstrate how Galaxy can support accessible, reproducible, and transparent NGS re-sequencing studies, we performed a simple variant analysis experiment. This experiment identifies variants from a set of 4,536,964 RNA-seq reads obtained from sequencing a sample of mm9 brain tissue from day 18 of embryonic development.

The initial analysis produced support for 27,742 possible variants. Of these possible variants, there are 5,625 where (a) the consensus base--as determined by the MAQ model--differs from the reference base and (b) read coverage at the base is 30x or greater. Of these potential variants, 2796 occur in known RefSeq Genes. These potential variants are:

Embedded Galaxy Dataset 'Variants from sample E18, consensus different, in RefSeq Genes'
[Do not edit this block; Galaxy will fill it in with the annotated dataset when it is displayed.]

Method

In the first step of this analysis, the reads were groomed to convert their quality scores from Solexa 1.0 to Solexa 1.3/Fastqsanger. Next, the reads were trimmed from 36bp to 27bp to exclude base pairs with low quality scores; see [1] for this step's rationale and parameter choices. After grooming and trimming, the reads were mapped using the short-read mapper Bowtie [2]. A pileup analysis using SAMtools [3] was then performed and was filtered to identify variants supported by 30+ reads. The complete analysis is contained in this history:

Embedded Galaxy History 'Variant Pileup Analysis for Sample E18'
[Do not edit this block; Galaxy will fill it in with the annotated history when it is displayed.]

Here is a workflow for performing this analysis:

Embedded Galaxy Workflow 'SNP identification within annotated genes from NGS PE Data'
[Do not edit this block; Galaxy will fill it in with the annotated workflow when it is displayed.]

References

[1] Han, X. et al. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proceedings of the National Academy of Sciences* 106, 12741-12746 (2009).

[2] Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

[3] Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).

The power of Galaxy publishing

Galaxy's publishing features facilitate access and reproducibility without any extra leg work

One click grants access to the *actual analysis* you performed to generate your original results

- ✦ Not just data access: the full pipeline
- ✦ Annotate each step
- ✦ Anyone can import your work and immediately reproduce or build on it

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ **public website**
- ✦ local instance
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Galaxy main site (<http://usegalaxy.org>)

Public web site, anybody can use

~500 new users per month, ~100 TB of user data,
~130,000 analysis jobs per month, every month is
our busiest month ever...

Will continue to be maintained and enhanced, but
with limits and quotas

Centralized solution cannot scale to meet data
analysis demands

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ public website
- ✦ **local instance**
- ✦ on the Cloud
- ✦ tool shed/contributing tools

Local Galaxy instances

(<http://getgalaxy.org>)

Galaxy is designed for local installation and customization

- ✦ Just download and run, completely self-contained
- ✦ Easily integrate new tools
- ✦ Easy to deploy and manage on nearly any (unix) system
- ✦ Run jobs on existing compute clusters

Especially useful for sensitive data

- ✦ can secure data and abide by regulations

Scale up on existing resources

Move intensive processing (tool execution) to other hosts



Frees up the application server to serve requests and manage jobs



Utilize existing resources



Supports any scheduler that supports DRMAA (most of them)



Running a **Production** Server

Use a real database server: PostgreSQL, MySQL

Run on compute cluster resources

External Authentication: LDAP, Kerberos, OpenID

Load balancing; proxy support

Lack IT knowledge or resources?

No problem, just use the **Cloud**

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ **on the Cloud**
- ✦ tool shed/contributing tools

Deploying Galaxy on the AWS Cloud

<http://usegalaxy.org/cloud>

1. **Open an AWS account** (only once)
2. Use the AWS Management Console to **start a master EC2 instance**
3. **Use the Galaxy CloudMan web interface** on the master instance to manage the cluster

2. Start an EC2 Instance

The screenshot displays the AWS Management Console interface. At the top, the navigation bar includes 'AWS', 'Products', 'Developers', 'Community', 'Support', and 'Account'. The 'Account' menu is highlighted with a red box. Below the navigation bar, the 'Your Account' section is visible, with 'Security Credentials' highlighted by a red box. The main content area shows the 'Amazon EC2 Console Dashboard' for the 'US East' region. A 'Launch Instance' button is prominently displayed. The 'Request Instances Wizard' is open, showing the following configuration details:

- AMI:** Other Linux AMI ID ami-ed03ed84 (x86_64) Edit AMI
- Number of Instances:** 1
- Availability Zone:** No Preference
- Monitoring:** Disabled
- Instance Type:** Large (m1.large)
- Instance Class:** On Demand Edit Instance Details
- Kernel ID:** Use Default
- Ramdisk ID:** Use Default
- User Data:** testGC1|AKIAJKQI3RT... Edit Advanced Details
- Key Pair Name:** galaxy_keypair Edit Key Pair
- Security Group(s):** default, galaxyWeb Edit Firewall

At the bottom of the wizard, there are 'Back' and 'Launch' buttons.

3. Configure Your Cluster

The screenshot shows a web browser window with the URL `ec2-50-16-1-149.compute-1.amazonaws.com/cloud`. The page title is "Galaxy Cloudman" and it includes navigation links for "report bugs", "wiki", and "screencast". The main content area is partially obscured by a modal dialog box titled "Initial Cluster Configuration".

Initial Cluster Configuration

Welcome to Galaxy Cloudman. This application will allow you to manage this cluster and the services provided within. To get started, choose the type of cluster you'd like to work with and specify the size of your persistent data storage, if any.

Start a full Galaxy Cluster. Specify initial storage size (in Gigabytes)

GB **OK**

[Show more startup options](#)

The background interface shows a "Status" section with fields for "Cluster name", "Disk status:", "Worker status:", "Service status:", and "External Logs". A "Cluster status log" section is visible at the bottom with a green plus icon.

Galaxy Cloudman Console

Welcome to Galaxy Cloudman. This application will allow you to manage this cloud and the services provided within. If this is your first time running this cluster, you will need to select an initial data volume size. Once the data store is configured, default services will start and you will be able to add and remove additional services as well as 'worker' nodes on which jobs are run.

[Terminate cluster](#) [Add nodes ▼](#) [Remove nodes](#) [Access Galaxy](#)

Status

Cluster name: ttt
Disk status: 0 / 0 (0%)
Worker status: Idle: 0 Available: 0 Requested: 0
Service status: Applications ● Data ●

- Pending
- Starting
- Ready
- Error

Cluster status log +

Tools

- [Get Data](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Operate on Genomic Intervals](#)
- [Graph/Display Data](#)

- NGS TOOLBOX BETA
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: SAM Tools](#)

Welcome to Galaxy on the Cloud

History

Options

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

The image displays two overlapping browser windows. The background window shows the Galaxy interface with a 'Saved Histories' table. The foreground window shows the 'Galaxy Cloud Console' with a 'Scale' section and a 'Cluster status log'.

Saved Histories Table:

Name	Datasets (by state)	Tags	Sharing	Created	Last Updated
mt_replicates_pair 2	8	96	0 Tags	about 1 hour ago	2 m ago
mt_replicates_pair 2	8	96	0 Tags	about 1 hour ago	15 min ago
mt_replicates_pair 1_testing	35	3	66	0 Tags	about 2 hours ago
mt_datasets	24		0 Tags	about 2 hours ago	abo

Galaxy Cloud Console:

Scale: Add more instances, Remove idle instances

Status: Cluster name: james-galaxy-cluster-9May2010-1, Cluster status: Ready, Instance status: Idle: 0 Available: 4 Requested: 4

Cluster status log:

```

14:54:40 - Instance 'i-a3e7b2c8' ready
14:54:40 - Setting up Galaxy
14:54:40 - Starting Galaxy...
14:54:45 - Instance 'i-a1e7b2ca' ready
14:54:49 - Instance 'i-afe7b2c4' ready
14:54:56 - Instance 'i-a3e7b2c8' reported alive
14:54:56 - Sent master public key to worker instance 'i-a3e7b2c8'.
14:55:00 - Adding instance 'i-a3e7b2c8' to SGE Execution Host list
14:55:01 - Successfully added instance 'i-a3e7b2c8' to SGE
14:55:01 - Waiting on worker instance 'i-a3e7b2c8' to configure itself...
14:55:09 - Instance 'i-a3e7b2c8' ready
14:55:16 - Galaxy started successfully!
14:55:16 - Ready for use

```

Can use like any other Galaxy instance, with additional compute nodes acquired and released (*automatically*) in response to usage

Overview

What is Galaxy?

What **you** can do in Galaxy

- ✦ analysis interface, tools and datasources
- ✦ data libraries
- ✦ workflows
- ✦ visualization
- ✦ sharing
- ✦ Pages

Where **you** can use and build Galaxy

- ✦ public website
- ✦ local instance
- ✦ on the Cloud
- ✦ **tool shed/contributing tools**

The Problem

You have written a Perl script to analyze genomic data and you want to share it with command-line averse colleagues

The Galaxy Solution

Solution: Integrate the script as a new Tool into your own Galaxy server

Steps:

- ✦ Obtain and install Galaxy source code (GetGalaxy.org)
- ✦ Write an XML file describing the inputs and outputs and how to execute the script
- ✦ Instruct Galaxy to load the tool

Adding your Own

Write or download a command-line executable

Determine number and kind of

- ✦ Input and Output Datasets
- ✦ Input Parameters

Construct a descriptive tool configuration XML file

- ✦ Write a wrapper script, only if required

Cluster

Cluster intervals of:

max distance between intervals: (bp)

min number of intervals per cluster:

Return type:

TIP: If your query does not appear in the pulldown menu -> it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns

Screencasts!

See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

Syntax

- **Maximum distance** is greatest distance in base pairs allowed between intervals that will be considered "clustered". **Negative** values for distance are allowed, and are useful for clustering intervals that overlap.
- **Minimum intervals per cluster** allow a threshold to be set on the minimum number of intervals to be considered a cluster. Any area with less than this minimum will not be included in the output.
- **Merge clusters into single intervals** outputs intervals that span the entire cluster.
- **Find cluster intervals; preserve comments and order** filters out non-cluster intervals while maintaining the original ordering and comments in the file.
- **Find cluster intervals; output grouped by clusters** filters out non-cluster intervals, but outputs the cluster intervals so that they are grouped together. Comments and original ordering in the file are lost.

Example



```
cluster.xml
1 <tool id="gops_cluster_1" name="Cluster">
2   <description>[[Cluster]] the intervals of a query</description>
3   <command interpreter="python2.4">
4     gops_cluster.py $input1 $output -l $input1_chromCol,$input1_startC
5       -d $distance -m $minregions -o $returntype
6   </command>
7   <inputs>
8     <param format="interval" name="input1" type="data">
9       <label>Cluster intervals of</label>
10    </param>
11    <param name="distance" size="5" type="integer" value="1" help="(bp
12      <label>max distance between intervals</label>
13    </param>
14    <param name="minregions" size="5" type="integer" value="2">
15      <label>min number of intervals per cluster</label>
16    </param>
17    <param name="returntype" type="select" label="Return type">
18      <option value="1">Merge clusters into single intervals</option>
19      <option value="2">Find cluster intervals; preserve comments and
20      <option value="3">Find cluster intervals; output grouped by clus
21      <option value="4">Find the smallest interval in each cluster</op
22      <option value="5">Find the largest interval in each cluster</opt
23    </param>
24  </inputs>
25  <help>
26
27  .. class:: infomark
28
29  **TIP:** If your query does not appear in the pulldown menu -> it is n
30
31  -----
32
33  **Screencasts!**
34
35  See Galaxy Interval Operation Screencasts (right click to open this l
36
37  .. \_Screencasts: http://www.bx.psu.edu/cgi-bin/trac.cgi/wiki/GopsDesc
38
39  -----
40
41  **Syntax**
42
43  - Maximum distance is greatest distance in base pairs allowed betw
44  - Minimum intervals per cluster allow a threshold to be set on the
45  - Merge clusters into single intervals outputs intervals that span
46  - Find cluster intervals; preserve comments and order filters out
47  - Find cluster intervals; output grouped by clusters filters out n
48
49  Line: 87 Column: 8 XML Soft Tabs: 2
```

Adding your Own Display Application

Define An XML configuration which describes how and where to present the data to the External Web Application




- ✦ Static
- ✦ Dynamic - display options can be loaded from a file

Inform Galaxy about the new display by adding to the appropriate datatype in `datatypes_conf.xml`

Static External Display Application

```
<display id="ucsc_bam" version="1.0.0" name="display at UCSC">  
  <link id="main" name="main">  
    <url>http://genome.ucsc.edu/cgi-bin/hgTracks?db=${qp($bam_file.dbkey)}&hgt.customText=${qp($track.url)}</url>  
    <param type="data" name="bam_file" url="galaxy.bam" strip_https="True" />  
    <param type="data" name="bai_file" url="galaxy.bam.bai" metadata="bam_index" strip_https="True" />  
    <param type="template" name="track" viewable="True" strip_https="True">  
      track type=bam name="${bam_file.name}" bigDataUrl=${bam_file.url} db=${bam_file.dbkey}  
    </param>  
  </link>  
</display>
```

```
<datatype extension="bam" type="galaxy.datatypes.binary:Bam"  
  mimetype="application/octet-stream" display_in_upload="true">  
  <display file="ucsc/bam.xml" />  
</datatype>
```

2: SAM-to-BAM on data 1   

660.5 Mb, format: bam, database:
mm9

Info:



| display at UCSC [main](#)

Binary bam alignments file

BAM at UCSC

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl PDF/PS Session Help

UCSC Genome Browser on Mouse July 2007 (NCBI37/mm9) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr12:57,795,963-57,815,592 gene jump clear size 19,630 bp. configure

chr12 (qC1) 12qA1.1 qA2 12qA3 qB1 12qB3 12qC1 12qC2 12qC3 qD1 qE 12qD3 12qE 12qF1 qF2

Scale 5 kb

to-BAM on data 1 SAM-to-BAM on data 1 STS Markers on Gene5.1c and Rad141c Hybrid Maps

UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

RefSeq Genes Non-House RefSeq Genes

Ensembl Gene Predictions

Human Proteins Mapped by Chained tBLASTn

Mouse mRNAs from GenBank

Spliced ESTs

Mouse ESTs That Have Been Spliced

36-way Multiz Alignment & Conservation

Maternal Cons

Rat Human Orangutan Dog Horse Opossum Chicken Stickleback

Simple Nucleotide Polymorphisms (dbSNP build 126)

RepeatMasker

move start < 2.0 > Click on a feature for details. Click or drag in the base position track to zoom in. move end < 2.0 >

Click gray/blue bars on left for track options and descriptions.

default tracks hide all manage custom tracks configure reverse refresh

Use drop-down controls below and press refresh to alter tracks displayed.

Tracks with lots of items will automatically be displayed in more compact modes.

collapse all expand all

Dynamic External Display Application

```
<display id="ucsc_bam" version="1.0.0" name="display at UCSC">
  <!-- Load links from file: one line to one link -->
  <dynamic_links from_file="tool-data/shared/ucsc/ucsc_build_sites.txt" skip_startswith="#" id="0" name="0">




    <!-- Define parameters by column from file, allow splitting on builds -->
    <dynamic_param name="site_id" value="0"/>
    <dynamic_param name="ucsc_link" value="1"/>
    <dynamic_param name="builds" value="2" split="True" separator="," />

    <!-- Filter out some of the links based upon matching site_id to a Galaxy application configuration parameter and b
    <filter>${site_id in $APP.config.ucsc_display_sites}</filter>
    <filter>${dataset.dbkey in $builds}</filter>

    <!-- We define url and params as normal, but values defined in dynamic_param are available by specified name -->
    <url>${ucsc_link}db=${qp($bam_file.dbkey)}&hgt.customText=${qp($track.url)}</url>
    <param type="data" name="bam_file" url="galaxy_${DATASET_HASH}.bam" strip_https="True" />
    <param type="data" name="bai_file" url="galaxy_${DATASET_HASH}.bam.bai" metadata="bam_index" strip_https="True" />
    <param type="template" name="track" viewable="True" strip_https="True">
      track type=bam name="${bam_file.name}" bigDataUrl=${bam_file.url} db=${bam_file.dbkey}
    </param>

  </dynamic_links>
</display>
```

```
#Harvested from http://genome.ucsc.edu/cgi-bin/das/dsn
main http://genome.ucsc.edu/cgi-bin/hgTracks? anoCar1,ce6,ce4,ce2,rn3,l
#Harvested from http://archaea.ucsc.edu/cgi-bin/das/dsn
archaea http://archaea.ucsc.edu/cgi-bin/hgTracks? therSibi1,symbTher_IAM148
#Harvested from http://main.genome-browser.bx.psu.edu/cgi-bin/das/dsn
bx-main http://main.genome-browser.bx.psu.edu/cgi-bin/hgTracks? oviAri1,eriEu
```

2: SAM-to-BAM on data 1   

660.5 Mb, format: bam, database: mm9

Info:



| display at UCSC [main](#) [bx-main](#)

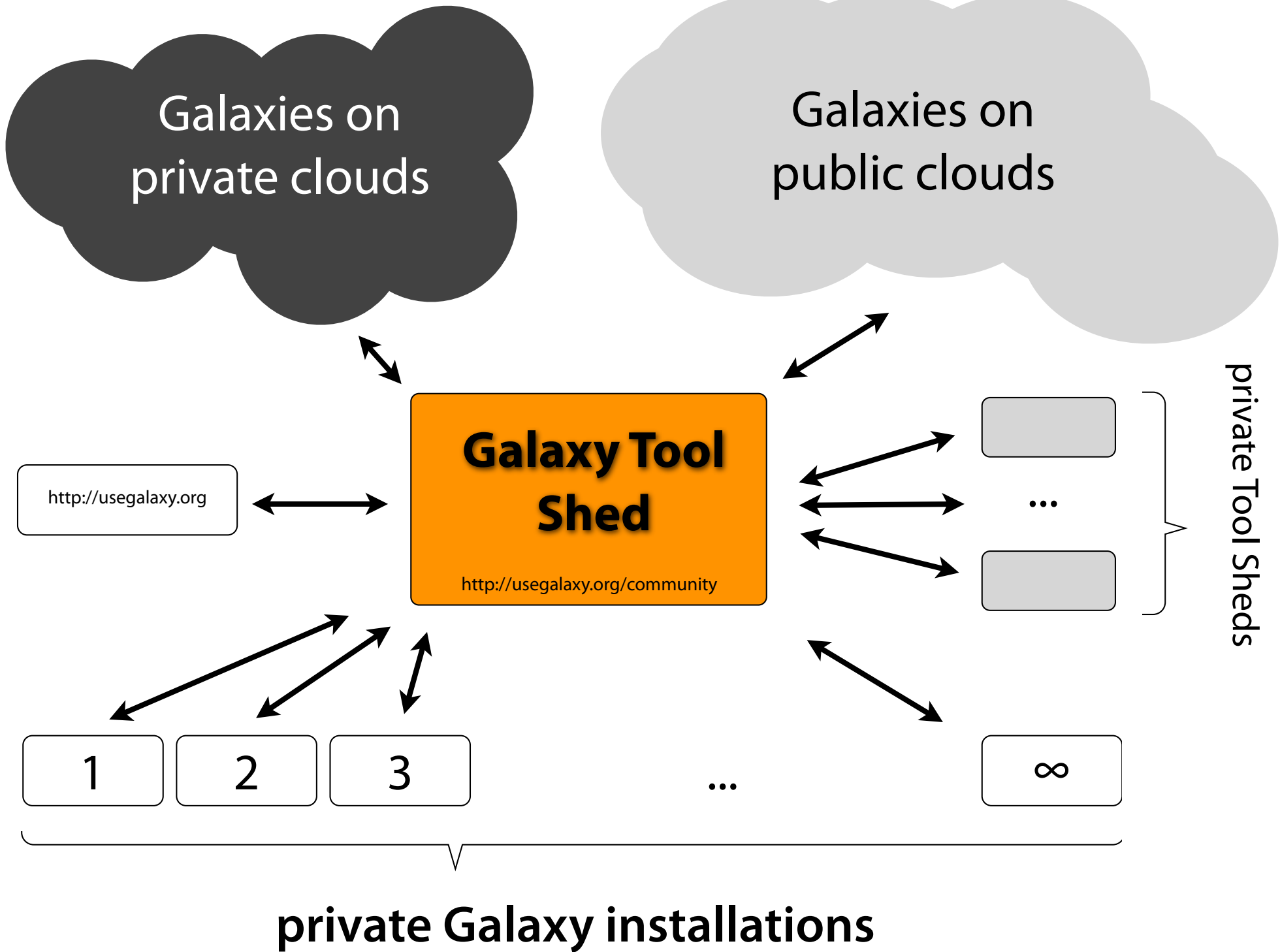
Binary bam alignments file

You added a tool, now what?

Share it with the community!

Galaxy Tool Shed

- ✦ Upload and Download contributed tools
- ✦ Rate and provide comments and feedback



Get and Contribute Tools


Galaxy Tool Shed / (beta) Tools Help User

Community

Tools

- [Browse by category](#)
- [Browse all tools](#)
- [Login to upload](#)

Categories

 [Advanced Search](#)

Name ↓	Description	Tools
Convert Formats	Tools for converting data formats	4
Data Source	Tools for retrieving data from external data sources	1
Fasta Manipulation	Tools for manipulating fasta data	5
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing data	5
Ontology Manipulation	Tools for manipulating ontologies	1
SAM	Tools for manipulating alignments in the SAM format	0
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	7
SNP Analysis	Tools for single nucleotide polymorphism data such as WGA	1
Statistics	Tools for generating statistics	1
Text Manipulation	Tools for manipulating data	3
Visualization	Tools for visualizing data	1

<http://usegalaxy.org/community>

Some future challenges

- Capturing and automatically deploying tool dependencies, automatic tool acquisition in Galaxy instances
- Better interfaces for highly parallel analysis (e.g. running the same workflow across 192 individuals)
- Various workflow engine improvements, partial data streaming, combined experimental/computational workflows

Try it now:

<http://usegalaxy.org>

Develop and deploy:

<http://getgalaxy.org>

<http://galaxyproject.org>

Come do cool stuff, contact us at:

[http://wiki.g2.bx.psu.edu/News/Galaxy is Hiring](http://wiki.g2.bx.psu.edu/News/Galaxy%20is%20Hiring)

Opportunities for collaboration, positions for postdocs, researchers, software engineers



EMORY

PENNSYLVANIA STATE UNIVERSITY



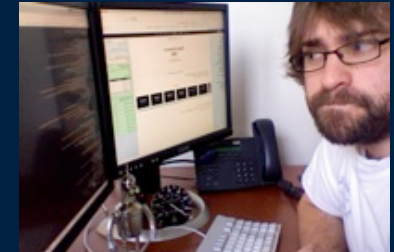
Enis Afgan



Dannon Baker



Dan Blankenberg



Nate Coraor



Dave Clements



Jeremy Goecks



Jennifer Jackson



Greg von Kuster



Kanwei Li



James Taylor



Guru Ananda



Anton Nekrutenko

Supported by the **NHGRI** (HG005542, HG004909, HG005133), **NSF** (DBI-0850103), Penn State University, Emory University, and the Pennsylvania Department of Public Health