

# Tutorial: Welcome to Galaxy

Research Computing Workshop 2011

Curtis Hendrickson (curtish@uab.edu)

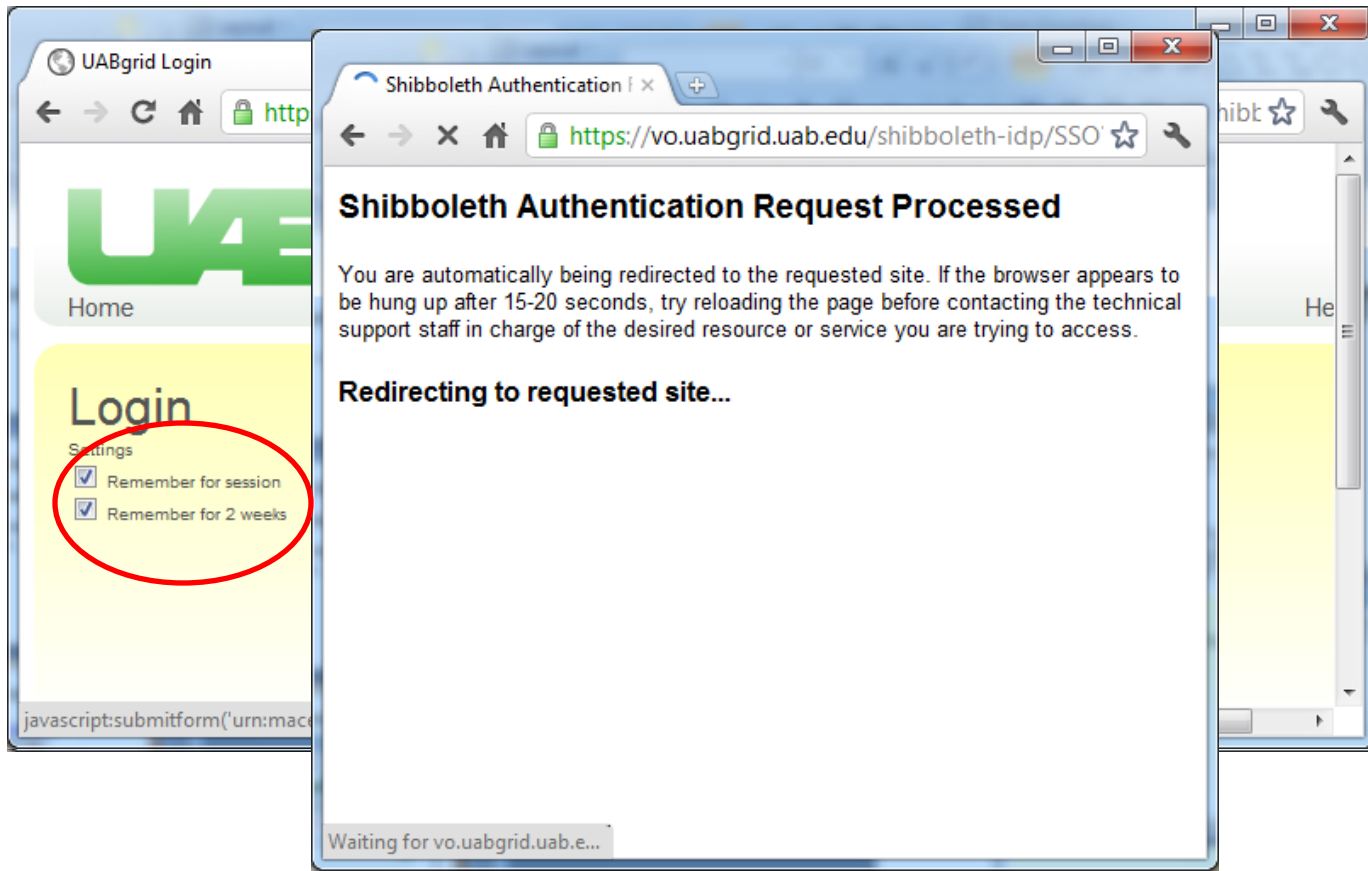
CCTS Biomedical Informatics Service

# What is Galaxy?

- Open source web application for data analysis
  - Command-line free cluster integration
  - Multi-**Tera**byte “scratch” storage in UAB cloud
- Galaxy manages data and workflows/protocol
  - Allows sharing of data & protocols
- Plug-ins for each tool, can add your own tools
  
- Today, we will discuss using it for Next Generation Sequence analysis
  - Note that EMBOSS sequence analysis tools installed

# First Time Login

- <https://galaxy.uabgrid.uab.edu>
- Re-direction to BlazerID login



# Homepage

Tools

History

**Galaxy / UAB** Analyze Data Workflow Shared Data Admin Help User Using 680.5 Gb

**Tools** Options ▾

- [Get Data](#)
- [Send Data](#)
- [ENCODE Tools](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Convert Formats](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Wavelet Analysis](#)
- [Graph/Display Data](#)
- [Regional Variation](#)
- [Multiple regression](#)
- [Multivariate Analysis](#)
- [Evolution](#)
- [Motif Tools](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [FASTA manipulation](#)
- [NCBI BLAST+](#)
- NGS TOOLBOX BETA
- [NGS: QC and manipulation](#)
- [NGS: Assembly](#)
- [NGS: Mannin](#)

**Welcome to UAB Galaxy!**

Welcome to the UAB Galaxy platform for experimental biology and comparative genomics designed to help you analyze multiple alignments, compare genomic annotations, profile metagenomic samples and more from your web browser. This platform is built on [Galaxy](#), backed by the [Cheaha compute cluster](#), and powered by [UABgrid](#). Documentation on the UAB installation can be found on the [UAB Galaxy wiki](#).

The UAB instance of Galaxy is live as of May 27th, 2011. Please be aware, however, that not all tools or data sets are currently available. Additional tools and data sets are planned, and more can be requested.

**Galaxy User Support:** In order to facilitate interaction among UAB Galaxy users, share experience, and provide peer-support we have established a galaxy-users group. To join this group and participate in email discussions please subscribe to the [galaxy-user](#) group. On-line archives of these discussions are available [here](#). Please note, the email discussions are a public forum. You are advised to only post information you are authorized to share and comfortable with being public.

Galaxy is developed by Penn State and Emory University. The UAB Galaxy platform is a collaborative project between the Biomedical Informatics group of the Center for Clinical and Translational Science and UAB IT Research Computing. This project is supported in part by the UAB Center for Clinical and Translational Science under grant UL1 RR025777 from the NIH National Center for Research Resources and by the Office of the Vice President for Information Technology at UAB. Please reference these in any publications resulting from your use of this platform.

**History** Options ▾

Unnamed history 0 bytes

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

WWFSMD?

# 1. Name your history

The screenshot displays the Galaxy / UAB web interface. At the top, a dark navigation bar contains the text "Galaxy / UAB" on the left, a menu of options including "Analyze Data", "Workflow", "Shared Data", "Admin", "Help", and "User" in the center, and a status indicator "Using 680.5 Gb" on the right. Below this, the interface is divided into three main vertical sections. The leftmost section is a "Tools" sidebar with a scrollable list of tool categories such as "Get Data", "Send Data", "ENCODE Tools", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Convert Formats", "Extract Features", "Fetch Sequences", "Get Genomic Scores", "Operate on Genomic Intervals", "Statistics", "Wavelet Analysis", "Graph/Display Data", "Regional Variation", "Multiple regression", "Multivariate Analysis", "Evolution", "Motif Tools", "Multiple Alignments", "Metagenomic analyses", "FASTA manipulation", "NCBI BLAST+", "NGS TOOLBOX BETA", "NGS: QC and manipulation", "NGS: Assembly", "NGS: Mapping", and "NGS: Total Analysis". The middle section is a large green-bordered box containing a "Welcome to UAB Galaxy!" message with a green checkmark icon. The text in this box includes a welcome message, information about the platform's backing by the Cheaha compute cluster and UABgrid, a note about the UAB instance being live as of May 27th, 2011, and a "Galaxy User Support" section encouraging users to join a support group. At the bottom of this box, it says "WWFSMD? grow noodly appendages". The rightmost section is a "History" panel with an "Options" dropdown. It shows a single history item named "RCD 2011" with a "Click to rename history" button and a "0 bytes" size indicator. Below the history item, a blue information box states: "Your history is empty. Click 'Get Data' on the left pane to start".

# Get Data – Upload File

The screenshot shows the Galaxy UAB web interface. The browser address bar displays <https://galaxy.uabgrid.uab.edu>. The main navigation bar includes 'Galaxy / UAB', 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', 'User', and 'Using 680.5 Gb'. The left sidebar contains a 'Tools' section with 'Options' and a list of tools. The 'Get Data' tool is circled in red. The main content area shows the 'Upload File' tool interface, which includes a 'File Format' dropdown set to 'Auto-detect', a 'File:' section with a 'Choose File' button and 'No file chosen' text, and a 'URL/Text:' section with a text input area. Below this is a table of files uploaded via FTP.

**Tools** Options

- Get Data**
- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX main browser
- Get Microbial Data
- BioMart Central server
- BioMart Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- metabolicMine server
- modENCODE worm server
- Wormbase server

**Upload File**

**File Format:**  
Auto-detect  
Which format? See help below

**File:**  
Choose File No file chosen  
TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

**URL/Text:**

Here you may specify a list of URLs (one per line) or paste the contents of a file.

**Files uploaded via FTP:**

File	Size
<input type="checkbox"/> C02KNABXX_s1_1_illumina12index_11_SL5780.fastq	3.5 Gb
<input type="checkbox"/> head.fastq	424 bytes
<input type="checkbox"/> C02KNABXX_s1_2_illumina12index_9-11.fastq	11.4 Gb

**History** Options

RCD 2011 0 bytes

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

# Upload File from Desktop

## set format & genome!

The screenshot shows the Galaxy UAB web interface. The browser address bar displays <https://galaxy.uabgrid.uab.edu>. The main navigation bar includes 'Galaxy / UAB', 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', 'User', and 'Using 680.5 Gb'. The left sidebar contains a 'Tools' section with an 'Options' dropdown and a 'Get Data' section with various data sources. The central panel shows the 'Upload File' tool configuration. The 'File Format' dropdown is set to 'fastqs', with 'fastqsanger' and 'fastqsolexa' visible in the dropdown menu. A red circle highlights the 'Choose File' button and the file name 'control\_mm9\_...ard.fastq.gz'. Below the file name, a tip states: 'TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator)'. The 'URL/Text' section has an empty text area. The 'Convert spaces to tabs' section has an unchecked checkbox. The 'Genome' dropdown is set to 'mm9', with 'Mouse July 2007 (NCBI37/mm9) (mm9)' and 'mm9 based HMOX1 null mutant with Z82244 (construct)' visible in the dropdown menu. The right sidebar shows a 'History' section with an 'Options' dropdown and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

Galaxy / UAB

Analyze Data Workflow Shared Data Admin Help User Using 680.5 Gb

Tools Options

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX main browser
- Get Microbial Data
- BioMart Central server
- BioMart Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- metabolicMine server
- modENCODE worm server
- Wormbase server

Upload File

File Format:

fastqs

fastqsanger

fastqsolexa

Choose File control\_mm9\_...ard.fastq.gz

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Convert spaces to tabs:

Yes

Use this option if you are entering intervals by hand.

Genome:

mm9

Mouse July 2007 (NCBI37/mm9) (mm9)

mm9 based HMOX1 null mutant with Z82244 (construct)

History Options

RCD 2011 0 bytes

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

# Upload File from Desktop

transfer and decompress

The screenshot shows the Galaxy UAB web interface. The browser address bar displays <https://galaxy.uabgrid.uab.edu>. The main navigation bar includes "Galaxy / UAB" and menu items: "Analyze Data", "Workflow", "Shared Data", "Admin", "Help", "User", and a status indicator "Using 680.5 Gb".

The left sidebar is titled "Tools" and contains a "Get Data" section with the following options:

- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX main browser
- Get Microbial Data
- BioMart Central server
- BioMart Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- metabolicMine server
- modENCODE worm server
- Wormbase server

The central workspace displays a blue information box with the following text:

**i** Your upload has been queued. History entries that are still uploading will be blue, and turn green upon completion.

**Please do not use your browser's "stop" or "reload" buttons until the upload is complete, or it may be interrupted.**

You may safely continue to use Galaxy while the upload is in progress. Using "stop" and "reload" on pages other than Galaxy is also safe.

The right sidebar is titled "History" and shows a list of jobs. The current job is highlighted in green:

**1:** [control mm9\\_chr15 Plekhh2-PiqF forward.fastq](#) 1.5 Mb



# 3. Get Data – Shared Data Libraries

The screenshot shows the Galaxy UAB interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', and 'User'. The 'Shared Data' menu is open, showing options like 'Data Libraries', 'Published Histories', 'Published Workflows', and 'Published Pages'. Below the navigation, a green banner displays a success message: '1 dataset imported into 1 history: RCD 2011'. The main content area shows a data library titled 'Tutorial Data Sets' with a table of datasets. The table has columns for 'Name', 'Uploaded By', 'Date', and 'File Size'. The second dataset, 'control\_mm9\_chr15\_Plekhh2-PiG reverse.fastq', is selected with a checked checkbox. Below the table, there is a 'For selected datasets:' section with a dropdown menu set to 'Import to current history' and a 'Go' button.

Name	Uploaded By	Date	File Size
control_mm9_chr15_Plekhh2-PiG forward.fastq	curtish@uab.edu	2011-09-13	1.5 Mb
<input checked="" type="checkbox"/> control_mm9_chr15_Plekhh2-PiG reverse.fastq	curtish@uab.edu	2011-09-13	1.5 Mb
drugged_mm9_chr15_Plekhh2-PiG forward.fastq	curtish@uab.edu	2011-09-13	1.4 Mb
drugged_mm9_chr15_Plekhh2-PiG reverse.fastq	curtish@uab.edu	2011-09-13	1.4 Mb

**TIP:** You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.

**TIP:** Several compression options are available for downloading multiple library datasets simultaneously:

- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats

# 4. Get Data – http/ftp server (NOT HudsonAlpha)

The screenshot shows the Galaxy UAB web interface. The browser address bar displays `https://galaxy.uabgrid.uab.edu/root`. The navigation menu includes 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', and 'User'. The 'Tools' sidebar on the left lists various data sources, with 'Get Data' circled in red. The 'Upload File' tool is active, and its 'URL/Text' section is also circled in red, containing the following text:

```
ftp://ftp.genome.uab.edu/ICS/Galaxy_
RNAseq_tutorial/drugged_mm9_chr15
_Plekhh2-PigF_forward.fastq.gz
ftp://ftp.genome.uab.edu/ICS/Galaxy
RNAseq_tutorial/drugged_mm9_chr15
```

The 'History' panel on the right shows a list of jobs, with the top three entries circled in green:

- 4: [ftp://ftp.genome.uab.edu/ICS/Galaxy\\_RNAseq\\_tutorial/drugged\\_mm9\\_chr15\\_Plekhh2-PigF\\_reverse.fastq](ftp://ftp.genome.uab.edu/ICS/Galaxy_RNAseq_tutorial/drugged_mm9_chr15_Plekhh2-PigF_reverse.fastq)
- 3: [ftp://ftp.genome.uab.edu/ICS/Galaxy\\_RNAseq\\_tutorial/drugged\\_mm9\\_chr15\\_Plekhh2-PigF\\_forward.fastq](ftp://ftp.genome.uab.edu/ICS/Galaxy_RNAseq_tutorial/drugged_mm9_chr15_Plekhh2-PigF_forward.fastq)
- 2: [control mm9 chr15 Plekhh2-PigF reverse.fastq](#)

The bottom entry in the history is [1: control mm9 chr15 Plekhh2-PigF forward.fastq](#).

# 5. Get Data – BIG FILES

move to UAB cloud

1. Get account on Cheaha cluster
2. Secure Transfer: scp, SSH Secure File Transfer <http://www.uab.edu/it/software>
3. Transfer to `/lustre/scratch/user/projects_name`
4. Uncompress (`/share/apps/galaxy/galaxy-tools/bin/q_gunzip`)
5. Get 10Gb network connection from IT!

The screenshot shows an SSH Secure File Transfer window with the following details:

- Local Name Table:**

Local Name	Size	Type	Modified
control_mm9_chr15_Plekh...	1,591,472	FASTQ File	09/13/2011 03:31:1...
control_mm9_chr15_Plekh...	1,564,769	FASTQ File	09/13/2011 03:31:1...
drugged_mm9_chr15_Plekh...	1,453,867	FASTQ File	09/13/2011 03:31:0...
drugged_mm9_chr15_Plekh...	1,430,199	FASTQ File	09/13/2011 03:31:0...
- Remote Name Table:**

Remote Name	Size	Type	Modified	Attribute
control_mm9_chr15_Plekh...	1,591,472	FASTQ File	09/13/2011 03:31:1...	-rw-r--r-
control_mm9_chr15_Plekh...	1,564,769	FASTQ File	09/13/2011 03:31:1...	-rw-r--r-
- Transfer Queue Table:**

Source File	Source Directory	Destination Directory	Size	Status	Speed	Time
drugged_mm9_chr15_...	\\genome-gendevel\mgbf\Curtis...	/lustre/scratch/curtish/RCD 201...	1,453,867	45%	88.4 kB/s	00:00:09
drugged_mm9_chr15_...	\\genome-gendevel\mgbf\Curtis...	/lustre/scratch/curtish/RCD 201...	1,430,199	Queued	0.0 kB/s	

Diagram: A blue circle labeled "Desktop" has a large blue arrow pointing to a blue cloud labeled "Cheaha".

# 5. Get Data – BIG FILES

Bring UAB cloud scratch files into Galaxy

The screenshot shows the Galaxy UAB web interface. The browser address bar displays `https://galaxy.uabgrid.uab.edu/library`. The navigation menu includes 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', and 'User'. The 'Using 680.5 Gb' status is visible in the top right.

The main content area is titled 'Upload files to a data library' and includes a 'Browse this data library' button. The 'Upload a directory of files' section contains the following form fields:

- Upload option:** A dropdown menu with 'Upload directory of files' selected.
- File Format:** A dropdown menu with 'fastqsanger' selected.
- Server Directory:** A dropdown menu with 'RCD 2011 Get Data' selected.
- Copy data into Galaxy?:** A dropdown menu with 'Copy files into Galaxy' selected.

Below the 'Copy data into Galaxy?' dropdown, there is explanatory text: 'Normally data uploaded with this tool is copied into Galaxy's configured "file\_path" location where Galaxy has a form of control over the data files. However, this may not be desired (especially for large NGS datasets), so using the option labeled "Link to files without copying into Galaxy" will force Galaxy to always read the data from its original path. Any symlinks encountered in the uploaded directory will be dereferenced once. That is, Galaxy will point directly to the file that is linked, but no other symlinks further down the line will be dereferenced.'

At the bottom, there is a 'Convert spaces to tabs:' section with an unchecked checkbox for 'Yes' and the instruction 'Use this option if you are entering intervals by hand.' Below that is a 'Genome:' section with a dropdown menu showing 'Mouse July 2007 (NCBI37/mm9) (m)'.

# Get Data – UCSC Annotation

The screenshot shows the Galaxy UAB web interface. The main content area is the UCSC Table Browser tool. The configuration is as follows:

- clade: Mammal
- genome: Mouse
- assembly: July 2007 (NCBI37/mm9)
- group: Genes and Gene Prediction Tracks
- track: UCSC Genes
- table: knownGene
- region: genome (selected), chr12:57795963-57815592
- identifiers (names/accessions): paste list (selected)
- filter: create
- intersection: create
- correlation: create
- output format: BED - browser extensible data
- output file: (blank)
- file type returned: plain text (selected)

A modal window titled "Paste In Identifiers for UCSC Genes" is open, showing a text input field containing "PigF". The window also includes a "submit" button and "clear" and "cancel" buttons. A blue arrow points from the "upload list" button in the main form to the modal window.

# Get Data – UCSC Annotation

The screenshot shows the Galaxy UAB web interface. The browser address bar displays <https://galaxy.uabgrid.uab.edu/root>. The navigation menu includes 'Home', 'Genomes', 'Genome Browser', 'Blat', 'Tables', 'Gene Sorter', 'PCR', 'Session', and 'FAQ'. The 'Tools' sidebar on the left lists various data sources under the 'Get Data' category, including 'UCSC Main table browser', 'UCSC Test table browser', 'UCSC Archaea table browser', 'BX main browser', 'Get Microbial Data', 'BioMart Central server', 'BioMart Test server', 'CBI Rice Mart rice mart', 'GrameneMart Central server', 'modENCODE fly server', 'Flymine server', 'Flymine test server', 'modENCODE modMine server', 'Ratmine server', 'YeastMine server', 'metabolicMine server', 'modENCODE worm server', 'Wormbase server', 'Wormbase test server', 'EuPathDB server', and 'EncodeDB at NHGRI'. The main content area is titled 'Output knownGene as BED' and contains the following configuration options:

- Include custom track header:**
  - name=
  - description=
  - visibility=
  - url=
- Create one BED record per:**
  - Whole Gene
  - Upstream by  bases
  - Exons plus  bases at each end
  - Introns plus  bases at each end
  - 5' UTR Exons
  - Coding Exons
  - 3' UTR Exons
  - Downstream by  bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

# Get Data – UCSC Annotation

The screenshot displays the Galaxy UAB web interface. The browser address bar shows the URL <https://galaxy.uabgrid.uab.edu/root>. The main navigation bar includes 'Galaxy / UAB' and menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', and 'User'. The top right corner indicates 'Using 680.5 Gb'.

**Tools** (Options):

- Get Data
  - Upload File from your computer
  - UCSC Main table browser
  - UCSC Test table browser
  - UCSC Archaea table browser
  - BX main browser
  - Get Microbial Data
  - BioMart Central server
  - BioMart Test server
  - CBI Rice Mart rice mart
  - GrameneMart Central server
  - modENCODE fly server
  - Flymine server
  - Flymine test server
  - modENCODE modMine server
  - Ratmine server
  - YeastMine server
  - metabolicMine server
  - modENCODE worm server
  - Wormbase server
  - Wormbase test server
  - EuPathDB server
  - EncodeDB at NHGRI

**Message:** A green box with a checkmark icon contains the following text:

The following job has been successfully added to the queue:

**5: UCSC Main on Mouse: knownGene (genome)**

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

**History** (Options):

RCD 2011 5.8 Mb

- 5: UCSC Main on Mouse: knownGene (genome)** (eye icon, refresh icon, delete icon)
- 4:** [ftp://ftp.genome.uab.edu/ICS/Galaxy\\_RNAseq\\_tutorial/drugged\\_mm9\\_chr15\\_Plekhh2-PigF\\_reverse.fastq](ftp://ftp.genome.uab.edu/ICS/Galaxy_RNAseq_tutorial/drugged_mm9_chr15_Plekhh2-PigF_reverse.fastq) (eye icon, refresh icon, delete icon)
- 3:** [ftp://ftp.genome.uab.edu/ICS/Galaxy\\_RNAseq\\_tutorial/drugged\\_mm9\\_chr15\\_Plekhh2-PigF\\_forward.fastq](ftp://ftp.genome.uab.edu/ICS/Galaxy_RNAseq_tutorial/drugged_mm9_chr15_Plekhh2-PigF_forward.fastq) (eye icon, refresh icon, delete icon)
- 2:** [control mm9\\_chr15\\_Plekhh2-PigF\\_reverse.fastq](#) (eye icon, refresh icon, delete icon)
- 1:** [control mm9\\_chr15\\_Plekhh2-PigF\\_forward.fastq](#) (eye icon, refresh icon, delete icon)

# Get Data – UCSC Annotation

1. Click Dataset “title” to see metadata
2. Click Eyeball to show data

The screenshot shows the Galaxy UAB web interface. The browser address bar displays `https://galaxy.uabgrid.uab.edu/root`. The main navigation bar includes "Galaxy / UAB", "Analyze Data", "Workflow", "Shared Data", "Admin", "Help", "User", and "Using 680.5 Gb".

On the left, the "Tools" panel is expanded to "Get Data", listing various data sources like "UCSC Main table browser", "BioMart Central server", etc.

The central panel displays a table of genomic coordinates:

chr17	87396598	87424741	uc008dun.1	0	-
87396793	87423232	0	6		
309,109,117,92,249,104, 0,11544,23142,24600,26406,28039,					

On the right, the "History" panel shows a list of jobs. The top job is "UCSC Main on Mouse: knownGene (genome)" with an eye icon circled in red. A red arrow points from this eye icon to the table in the center. Below it, the file path is shown: `ftp://ftp.genome.uab.edu/ICS/Galaxy_RNAseq_tutorial/druqqed_mm9_chr15_Plekhh2-PigF_reverse.fastq`.



# Get Data – UCSC Annotation

1. Click “pencil” to edit metadata
2. Click “disk” to download

The screenshot shows the Galaxy UAB web interface. The browser address bar displays <https://galaxy.uabgrid.uab.edu/root>. The main navigation bar includes 'Galaxy / UAB', 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', 'User', and 'Using 680.5 Gb'.

**Tools Panel (Left):** Under 'Get Data', several tools are listed, including 'UCSC Main table browser', 'UCSC Test table browser', 'UCSC Archaea table browser', 'BX main browser', 'Get Microbial Data', 'BioMart Central server', 'BioMart Test server', 'CBI Rice Mart rice mart', 'GrameneMart Central server', 'modENCODE fly server', 'Flymine server', 'Flymine test server', 'modENCODE modMine server', 'Ratmine server', 'YeastMine server', 'metabolicMine server', and 'modENCODE worm server'.

**Edit Attributes Panel (Center):** The 'Name' field is highlighted with a green circle and contains the text 'UCSC Main: [PigF](#)'. The 'Database/Build' field is set to 'Mouse July 2007 (NCBI37/mm9)'. The 'Chrom column' is set to '1', 'Start column' to '2', and 'End column' to '3'. The 'Annotation / Notes' field is empty.

**History Panel (Right):** The 'History' panel shows a list of datasets. The top entry is 'RCD 2011' (5.8 Mb). Below it is '5: UCSC Main on Mouse: knownGene (genome)' (1 region, format: bed, database: mm9). A red arrow points from the 'pencil' icon next to this entry to the 'Name' field in the 'Edit Attributes' panel. Below this entry is a table with columns '1.Chrom', '2.Start', '3.End', and '4.Name'. The table contains one row: 'chr17', '87396598', '87424741', and 'uc008dun'. Below the table is a 'disk' icon (highlighted with a red circle) and a 'download' icon. The bottom entry is '4: ftp://ftp.genome.uab.edu/ICS/Galaxy\_RNAseq\_tutorial/druqqed\_mm9\_chr15\_Plekhh2-PigF\_reverse.fastq'. The bottom entry is '3: ftp://ftp.genome.uab.edu/ICS/Galaxy\_RNAseq\_tutorial/druqqed\_mm9\_chr15\_Plekhh2-PigF\_forward.fastq'.

# Get Data – UCSC Annotation

## 1. Eyeball of large data shows a sample

The screenshot displays the Galaxy web interface. The main content area shows a large dataset with a yellow warning box that reads: "This dataset is large and only the first megabyte is shown below. Show all | Save". The dataset content is a text file with genomic coordinates and annotations, such as "@HWI-ST534\_129:2:41:11498:22644:ATCACG" and "GGGCTGAGAGCCGGGGAAGCCGCGGAGCCGGGTGACTGGCGAGCCGGGA".

The left sidebar contains a "Tools" panel with a "Get Data" section. The right sidebar contains a "History" panel with a list of datasets. The top dataset in the history is "RCD 2011" (5.8 Mb). Below it is a dataset titled "5: UCSC Main table browser" with a sub-entry "Mouse: knownGene (genome)". A red circle highlights the "eyeball" icon next to this entry, and a red arrow points from the warning box to this icon. Below that is another dataset titled "4: ftp://ftp.genome.uab.edu/ICS/Galaxy\_RNAseq\_tutorial/druqqed\_mm9\_chr15\_Plekhh2-PigF\_reverse.fastq" (1.4 Mb). The bottom dataset is "3: ftp://ftp.genome.uab.edu/ICS/Galaxy\_RNAseq\_tutorial/dru".