Using Galaxy for High-throughput Sequencing (HTS) Analysis

Jeremy Goecks
The Galaxy Team
http://usegalaxy.org
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
- Prepare, quality control and manipulate reads
- Read Mapping
- SNP & INDEL analysis
- Binding sites analysis and peak calling
- Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
HTS Data

From the Sequencer:
- reads and quality scores (FASTQ)

In the Analysis Pipeline / Workflow:
- alignments against reference genome (SAM, BAM)
- annotations (GFF, BED)
- genome Assemblies (FASTA)
- quantitative tracks, e.g. conservation (WIG)
FASTQ Quality Scores

@UNIQUE_SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAAATAGTAAATCCATTGTTCAACTCAGTTT
+
!"*((((+++))%%%%+++)(%%%%%).1+++--+++)*55CCF>>>>>>>>>CCCCCCC65

http://en.wikipedia.org/wiki/FASTQ_format

Galaxy tools generally use Sanger format
- Need to convert quality scores to Sanger using Groomer tool
Getting Your Data into Galaxy

Cannot upload any file larger than 2GB via Web browser
  * Galaxy does not currently support compressed files

Use FTP client, e.g. FileZilla: http://filezilla-project.org/
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
  • Prepare, quality control and manipulate reads
  • Read Mapping
  • SNP & INDEL analysis
  • Binding sites analysis and peak calling
  • Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
Prepare and Quality Check

**SOLiD**
- SOLiD2fastq

**Illumina**
- Groomer
  - Quality Statistics
  - Generic Box Plot Tool
  - Read Trimmer
  - Quality Filter

**454**
- low quality splitter
- 4542fastq

---
Combining Sequences and Qualities

This tool joins a FASTA file to a Quality Score file, creating a single FASTQ block for each read.

What it does

Specifying a set of quality scores is optional; when not provided, the output will be fastsanger or fastqc/sanger (when a csfasta is provided) with each quality score being the maximal allowed value (93).

Use this tool, for example, to convert 454-type output to FASTQ.

Galaxy Interface Screenshot
Grooming --> Sanger

FASTQ Groomer

File to groom:
3: Combine FASTA and... and data 2

Input FASTQ quality scores type:
- Sanger
- Solexa
- Illumina 1.3+
- Sanger
- Color Space Sanger

What it does
This tool offers several conversions options relating to the FASTQ format.

When using Basic options, the output will be Sanger formatted Sanger.

When converting, if a quality score falls outside of the target, the minimum or maximum.

When converting between Solexa and the other formats, quality scores may be transformed using the equations found in Cock PL, Fields CI, Goto N, Heuer ML, quality scores, and the Solexa/Illumina FASTQ variants. Nucleotides are written as 'N' when quality score space.

When converting between color space (csSanger) and base space, the base 'G' is used as the adapter. The adapter bases are lost or gained; if gained, the base 'G' is used as the adapter. You cannot convert a color space read to base space if there is no adapter present in the color space encoding. Any masked or ambiguous nucleotides in base space will be converted to 'N's when determining color space encoding.

Quality Score Comparison

<table>
<thead>
<tr>
<th></th>
<th>Sanger</th>
<th>Phred+33, 93 values</th>
<th>(0, 93) (0 to 60 expected in raw reads)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Solexa</td>
<td>Phred+64, 62 values</td>
<td>(0, 62) (0 to 40 expected in raw reads)</td>
</tr>
</tbody>
</table>

Diagram adapted from http://en.wikipedia.org/wiki/FASTQ_format
Quality Statistics and Box Plot Tool

NGS TOOLBOX BETA
NGS: QC and manipulation
ILLUMINA DATA
- FASTQ Groomer convert between various FASTQ quality formats
- FASTQ splitter on joined paired end reads
- FASTQ joiner on paired end reads
- FASTQ Summary Statistics by column

Graph/Display Data
- Histogram of a numeric column
- Scatterplot of two numeric columns
- Plotting tool for multiple series and graph types
- Boxplot of quality statistics
FastQC

FastQC Report

Mon 20 Jun 2011

dataset_1750787.dat

Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Kmer Content

Basic Statistics

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filename</td>
<td>dataset_1750787.dat</td>
</tr>
<tr>
<td>File type</td>
<td>Conventional base calls</td>
</tr>
<tr>
<td>Encoding</td>
<td>Sanger / Illumina 1.9</td>
</tr>
</tbody>
</table>
FASTQ Trimmer

**FASTQ File:**
2: imported: GM12878_chr Dataset

**Define Base Offsets as:**
- Absolute Values

Use Absolute for fixed length reads (Illumina, SOLiD)
Use Percentage for variable length reads (Roche/454)

**Offset from 5’ end:**
0
Values start at 0, increasing from the left

**Offset from 3’ end:**
16
Values start at 0, increasing from the right

**Keep reads with zero length:**
☐

**Execute**

FASTQ Quality Trimmer

**FASTQ File:**
7: FASTQ Trimmer on data 2

**Keep reads with zero length:**
☐

**Trim ends:**
5’ and 3’

**Window size:**
1

**Step Size:**
1

**Maximum number of bases to exclude from the window during aggregation:**
0

**Aggregate action for window:**
min score

**Trim until aggregate score is:**

>=

**Quality Score:**
0.0

**Execute**

This tool allows you to trim the ends of reads.

You can specify either absolute or percent-based offsets. When using the percent-based method, offset
will be a percentage of the read's length.

For example, if you have a read of length 36:

```bash
# Some FASTQ Sanger Read
CAATAGTGC72CAGTGAAGTGTAGA7NAGCA
+1@8:B-8786;C8A=70783CA4-488<;3<38
And you set absolute offsets of 2 and 0:
```
Quality Filtering

**FASTQ File:**
7: FASTQ Trimmer on data 2

Requires groomed data: if your data does not appear here try using the FASTQ groomer.

**Minimum Size:**
0

**Maximum Size:**
0

A maximum size less than 1 indicates no limit.

**Minimum Quality:**
0.0

**Maximum Quality:**
0.0

A maximum quality less than 1 indicates no limit.

**Maximum number of bases allowed outside of quality range:**
0

This is paired end data:

Quality Filter on a Range of Bases

Add new Quality Filter on a Range of Bases

Execute

**Quality Filter on a Range of Bases**

**Quality Filter on a Range of Bases 1**

**Define Base Offsets as:**

- Absolute Values

Use Absolute for fixed length reads (Illumina, SOLiD)
Use Percentage for variable length reads (Roche/454)

**Offset from 5' end:**
0

Values start at 0, increasing from the left

**Offset from 3' end:**
0

Values start at 0, increasing from the right

**Aggregate read score for specified range:**

- min score

**Keep read when aggregate score is:**

- >=

**Quality Score:**
0.0

Remove Quality Filter on a Range of Bases 1

Add new Quality Filter on a Range of Bases

Execute
Manipulate FASTQ

FASTQ File:
7: FASTQ Trimmer on data 2
Requires groomed data: if your data does not appear here try using the FASTQ groomer.

Match Reads
Add new Match Reads

Manipulate Reads
Add new Manipulate Reads
Execute

Match Reads
Match Reads by:
Sequence Content

Sequence Match Type:
Regular Expression

Match by:
N

Remove Match Reads 1

Add new Match Reads

Manipulate Reads
Add new Manipulate Reads
Execute

Match Reads
Match Reads by:
Sequence Content

Sequence Match Type:
Regular Expression

Match by:
N

Remove Match Reads 1

Add new Match Reads

Manipulate Reads
Add new Manipulate Reads
Execute

Match Reads
Match Reads by:
Sequence Content

Sequence Match Type:
Regular Expression

Match by:
N

Remove Match Reads 1

Add new Match Reads

Manipulate Reads
Add new Manipulate Reads
Execute
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
• Prepare, quality control and manipulate reads
• Read Mapping
• SNP & INDEL analysis
• Binding sites analysis and peak calling
• Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
Mapping HTS Data

Collection of interchangeable mappers
- accept fastq format, produce SAM/BAM

Mappers for
- DNA
- RNA
- Local realignment
Mappers

DNA
- short reads: Bowtie, BWA, BFAST, PerM
- longer reads: LASTZ

Metagenomics
- Megablast

RNA / gapped-reads mapper
- Tophat
Commonly Used/Default Parameters
Full Parameter List

LASTZ is a high performance pairwise sequence aligner derived from BLASTZ. It is written by Bob Harris in Webb Miller's laboratory at Penn State University. Special scoring sets were derived to improve runtime performance and quality. This Galaxy version of LASTZ is geared towards aligning short (Illumina/Solexa, AB/SOLiD) and medium (Roche/454) reads against a reference sequence. There is excellent, extensive documentation on LASTZ available here.

Do you want to modify the reference name?:
No

Do not report matches below this identity (%):
0

Do not report matches above this identity (%):
100

Do not report matches that cover less than this percentage of each read:
0

Convert lowercase bases to uppercase:
Yes

Input formats

LASTZ accepts reference and reads in FASTA format. However, because Galaxy supports implicit format conversion the tool will recognize fastq and other method specific formats.
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
- Prepare, quality control and manipulate reads
- Read Mapping
- SNP & INDEL analysis
- Binding sites analysis and peak calling
- Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
SNPs & INDELs

SNPs from Pileup
- Generate
- Filter

NGS: SAM Tools
- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files

Indel Analysis

Select sam file to analyze:

Frequency threshold: 0.015

What it does
Given an input sam file, this tool provides analysis of the indels. It filters out matches that do not meet the frequency threshold. The way this frequency of occurrence is calculated is different for deletions and insertions. The CIGAR string's "M" can indicate an exact match or a mismatch. For SAM containing the following bits of information (assuming the reference "ACTGCTCAGT"):

The following totals would be calculated (this is an intermediate step and not output):
GATK Tools

Local re-alignment
Base re-calibration
Genotyping

Alpha status
• please try, report bugs
• available on test server: http://test.g2.bx.psu.edu/
Unified Genotyper

**Inputs**
- BAM files

*Lots of possible parameters*

**Output**
- VCF file(s)
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
- Prepare, quality control and manipulate reads
- Read Mapping
- SNP & INDEL analysis
- Binding sites analysis and peak calling
- Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
Peak Calling / ChIP-seq analysis

Punctate binding
- transcription factors

Diffuse binding
- histone modifications
- PolII
Punctate Binding --> MACS

Inputs
✦ Enriched Tag file
✦ Control / Input file (optional)

Outputs
✦ Called Peaks
✦ Negative Peaks (when control provided)
✦ Shifted Tag counts (wig, convert to bigWig for visualization)

I have Peaks, now what?

A Intersect
First query
Intervals to intersect with (second query)
Overlapping intervals
Overlapping pieces of intervals

B Subtract
First query
Intervals to subtract (second query)
Intervals with no overlap
Non-overlapping pieces of intervals

C Merge
Query
Selected intervals

D Concatenate
First query
Second query
Concatenated

E Complement
Query
Complement

F Cluster
Query
Find clusters
Merge clusters

Compare to other annotations using interval operations
Secondary Analysis

A simple goal: determine number of peaks that overlap a) coding exons, b) 5-UTRs, c) 3-UTRs, d) introns and d) other regions

Get Data
- Import Peak Call data
- Retrieve Gene location data from external data resource
- Extract exon and intron data from Gene Data (Gene BED To Exon/Intron/Codon BED expander x4)
- Create an Identifier column for each exon type (Add column x4)
- Create a single file containing the 4 types (Concatenate)
  - Complement the exon/intron intervals
  - Force complemented file to match format of Gene BED expander output (convert to BED6)
- Create an Identifier column for the ‘other’ type (Add column)
  - Concatenate the exons/introns and other files
- Determine which Peaks overlap the region types (Join)
- Calculate counts for each region type (Group)
Secondary Analysis
Annotation Profiler

One click to determine base coverage of the interval (or set of intervals) by a set of features (tables) available from UCSC galGal3, mm8, panTro2, rn4, canFam2, hg18, hg19, mm9, rheMac2
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
- Prepare, quality control and manipulate reads
- Read Mapping
- SNP & INDEL analysis
- Binding sites analysis and peak calling
- Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
Transcriptome Analysis
(with a reference genome)

TopHat

Cufflinks/compare/diff

TopHat

Map RNA (FASTQ) to a reference Genome
• gapped mapper

Outputs
• BAM file of accepted hits
• BED file of splice junctions
Cufflinks

Goal: transcript assembly and quantitation

Input: aligned RNA-Seq reads, usually from TopHat

Outputs
- assembled transcripts (GTF)
- genes’ and transcripts’ coordinates, expression levels
Cuffcompare

Goals
- generate complete list of transcripts for a set of transcripts
- compare assembled transcripts to a reference annotation

Inputs: assembled transcripts from Cufflinks

Outputs:
- Transcripts Combined File
- Transcripts Accuracy File
- Transcripts Tracking Files
Cuffdiff

Goals
- differential expression testing
- transcript quantitation

Inputs
- Combined set of transcripts
- mapped reads from 2+ samples

Outputs
- differential expression tests for transcripts, genes, splicing, promoters, CDS
- quantitation values for most elements
Next Steps

Filtering
- for differentially expressed elements
- combined transcripts (e.g. for those differentially expressed between samples)

Extract transcript sequences and profile sequences for function
Integrating Tools and Visualization
Working to add GATK Unified Genotyper to Trackster as well
Working with HTS Tools

Often challenging
- many parameters
- time intensive
- evaluating results difficult

Good options
- filter early, filter often: easier to understand fewer results
- experimentation: can rerun tools, workflows
- visualization: use tools in Trackster when possible
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
- Prepare, quality control and manipulate reads
- Read Mapping
- SNP & INDEL analysis
- Binding sites analysis and peak calling
- Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
Sample information tracked in Galaxy, state changes through laboratory workflow are captured, data is linked back to sample in user’s workspace.
Sample Tracking System

Built-in system for tracking sequencing requests

Customizable interfaces
- Sequencing Facility Managers/Administrators
- Users/Biologists

Streamlines data delivery: sequencing runs to users
How does it all work?

Customer:
1. Create Galaxy sequencing request
2. Add samples
3. Submit request

Facility manager:
1. Transfer datasets to Galaxy data library
2. Update sample states
3. Assign barcodes to samples
Sequencing Facility Managers

Setup the Galaxy sample tracking system according to the core facility workflow [Once per request type]

Create and submit a sequencing request on behalf of another user

Reject an incomplete or erroneous sequencing request

Receive samples and assign them tracking barcodes.

Setup data transfer from the sequencer
Sequencing Facility Users

Create and submit a sequencing request
Edit and resubmit a rejected sequencing request
Obtain datasets at the end of a sequencing run
Select Libraries and Histories, and Workflows to populate and run on sequenced samples.
Configure Available Request / Sample Options

![Galaxy Administration Interface]

- **Forms**
  - **Name**
  - **Description**
  - **Type**
  - Options:
    - Analysis Portal run details
    - Atlantic Biosciences Analysis Portal Form
    - Atlantic Biosciences request
    - Atlantic Biosciences sample

- **Options**
  - Create new form
  - Search
  - Advanced Search

- **Actions**
  - For 0 selected forms: Delete, Undelete
Configurations can be
- custom-built
- loaded from provided configuration files
Configure the Sequencer
User Creates a Request

Sequencing Requests

Create a new sequencing request

Select a request type configuration:

Select one

Atlantic Biosciences
User Creates a Request

Sequencing Requests

Create a new sequencing request

Select a request type configuration:
- Atlantic Biosciences

Name of the Experiment
My first ChIP-seq Experiment
(Required)

Description
This is Experiment was performed using the protoc
(Optional)

Name

Scientific Contact
dan@bx.psu.edu office address

Save Add samples
User Adds a Sample

### Add Samples to Sequencing Request "My first ChIP-seq Experiment"

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>Data Library</th>
<th>Folder</th>
<th>History</th>
<th>Workflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample_1</td>
<td></td>
<td>Dan's Sequencing Requests</td>
<td>ChIP-seq</td>
<td>My own ChIP-seq Experiment1</td>
<td>Dan's ChIP-seq Workflow</td>
</tr>
</tbody>
</table>

For each sample, select the data library and folder in which you would like the run datasets deposited. To automatically run a workflow on run datasets, select a history first and then the desired workflow.

**Layout**

Copy **1** samples from sample **None**

Select the sample from which the new sample should be copied or leave selection as None to add a new "generic" sample.

- **Add sample**
- **Save**
- **Cancel**

Click the Add sample button for each new sample and click the Save button when you have finished adding samples.

**Import samples from csv file**
### Add Samples to Sequencing Request "My first ChIP-seq Experiment"

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>Data Library</th>
<th>Folder</th>
<th>History</th>
<th>Workflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample_1</td>
<td>Unsubmitted</td>
<td>Dan's Sequencing Requests</td>
<td>ChIP-seq</td>
<td>My own ChIP-seq Experiment!</td>
<td>Dan's ChIP-seq Workflow</td>
</tr>
</tbody>
</table>

For each sample, select the data library and folder in which you would like the run datasets deposited. To automatically run a workflow on run datasets, select a history first and then the desired workflow.

**Layout**

**Copy 1 samples from sample** None

Select the sample from which the new sample should be copied or leave selection as None to add a new "generic" sample.

**Add sample**

Click the Add sample button for each new sample.

**Import samples from csv file**
Samples enter “New” state

The sequencing request has been submitted.

Sequencing request "My first ChIP-seq Experiment"

Current state: In Progress

Description: This is Experiment was performed using the protocol ...

User: dan@bx.psu.edu

Request type: Atlantic Biosciences

More

<table>
<thead>
<tr>
<th>Samples</th>
<th>Barcode</th>
<th>State</th>
<th>Data Library</th>
<th>Folder</th>
<th>History</th>
<th>Workflow</th>
<th>Run Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample_1</td>
<td>New</td>
<td>Dan’s Sequencing Requests</td>
<td>ChIP-seq</td>
<td>My own ChIP-seq Experiment!</td>
<td>Dan’s ChIP-seq Workflow</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Layout1
sequencing facility is informed of request

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Samples</th>
<th>Type</th>
<th>Last Updated</th>
<th>State</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>My first ChIP-seq</td>
<td>Experiment was performed using the protocol...</td>
<td>1</td>
<td>Atlantic Biosciences</td>
<td>26 minutes ago</td>
<td>In Progress</td>
<td><a href="mailto:dan@bx.psu.edu">dan@bx.psu.edu</a></td>
</tr>
<tr>
<td>new request</td>
<td></td>
<td>1</td>
<td>Atlantic Biosciences</td>
<td>3 days ago</td>
<td>Complete</td>
<td><a href="mailto:customer@corp.com">customer@corp.com</a></td>
</tr>
<tr>
<td>some experiment</td>
<td>a test description</td>
<td>1</td>
<td>Atlantic Biosciences</td>
<td>3 days ago</td>
<td>Complete</td>
<td><a href="mailto:customer@corp.com">customer@corp.com</a></td>
</tr>
</tbody>
</table>
## Sequencing Facility Receives Samples

<table>
<thead>
<tr>
<th>Name</th>
<th>Barcode</th>
<th>State</th>
<th>Data Library</th>
<th>Folder</th>
<th>History</th>
<th>Workflow</th>
<th>Run Datasets</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample_1</td>
<td></td>
<td>New</td>
<td></td>
<td></td>
<td></td>
<td>My Own ChIP-seq</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For selected samples: **Select one**

For each sample, select the data library and folder in which you would like the run datasets deposited. To automatically run a workflow on run datasets, select a history first and then the desired workflow.

### Sequencing Requests

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Samples</th>
<th>Type</th>
<th>Last Updated</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>My first ChIP-seq Experiment</td>
<td>This is Experiment was performed using the protocol ...</td>
<td>1</td>
<td>Atlantic Biosciences</td>
<td>35 minutes ago</td>
<td>Complete</td>
</tr>
</tbody>
</table>

### Facility
- assigns a barcode to sample tubes
- Scans barcode at each step to change state

### User can watch progress of sequencing request
Sequencing Finished

Datasets are transferred from sequencer into Galaxy
- library
- user’s history

Galaxy Workflow is executed on Dataset

User is automatically emailed
Extending Sample Tracking with ngLims

An add-on written by community contributor Brad Chapman

http://bitbucket.org/chapmanb/galaxy-central

https://bitbucket.org/galaxy/galaxy-central/wiki/LIMS/nglims

Sample tracking is completely extensible

Track manually, with barcodes, or integrate with an existing LIMS

Everything is configuration driven, capture whatever data and support whatever workflow you want

Interaction with sequence instruments and secondary analysis is completely pluggable
  * For services that provide a web / REST API even easier
Example: extensions from Brad Chapman for flowcell layout, multiplexing, ...
Overview

High-throughput Sequencing (HTS) Data

Using Galaxy to Analyze HTS Data
- Prepare, quality control and manipulate reads
- Read Mapping
- SNP & INDEL analysis
- Binding sites analysis and peak calling
- Transcriptome analysis

Galaxy for Sequencing Facilities

Galaxy exercises: ChIP-seq and RNA-seq
Supported by the **NHGRI** (HG005542, HG004909, HG005133), **NSF** (DBI-0850103), Penn State University, Emory University, and the Pennsylvania Department of Public Health
Using Galaxy

Use public Galaxy server: UseGalaxy.org
Download Galaxy source: GetGalaxy.org
Galaxy Wiki: GalaxyProject.org
Screencasts: GalaxyCast.org

Public Mailing Lists
- galaxy-bugs@bx.psu.edu
- galaxy-user@bx.psu.edu
- galaxy-dev@bx.psu.edu
ChIP-seq and RNA-seq exercises

http://usegalaxy.org/u/jeremy/p/galaxy-rna-seq-analysis-exercise

- Shared Data --> Published Histories --> Import
- start Tophat mapping first (second section), then look at QC (first section)

http://usegalaxy.org/u/james/p/exercise-chip-seq

http://usegalaxy.org/u/jeremy/p/usc-exercise-clusters