How To: Run the ENCODE long-RNA-seq analysis pipeline on DNAnexus

Overview: In this exercise, we will run the ENCODE Uniform Processing Long RNA-seq Pipeline on a small test dataset containing reads from chromosome 21 sampled from an ENCODE RNA-seq experiment on a stomach tissue sample.

The ENCODE Portal page for the experiment is here: (https://www.encodeproject.org/experiments/ENCSR000AFI/)

The pipeline was specified by the ENCODE RNA Working Group and implemented at the ENCODE Data Coordinating Center (DCC). Today we will run the pipeline on the DNAnexus cloud platform. Typically, full ENCODE RNA experiments run on this pipeline are whole genome 30x read depth and take around 10 hours. This demonstration dataset can be processed in about 46 minutes.

The ENCODE pipeline code is open-source and lives on github at: https://github.com/ENCODE-DCC/long-rna-seq-pipeline. The pipeline is modeled on the encode portal which provides links directly to the exact scripts that define each step: https://www.encodeproject.org/pipelines/ENCPL002LPE/.

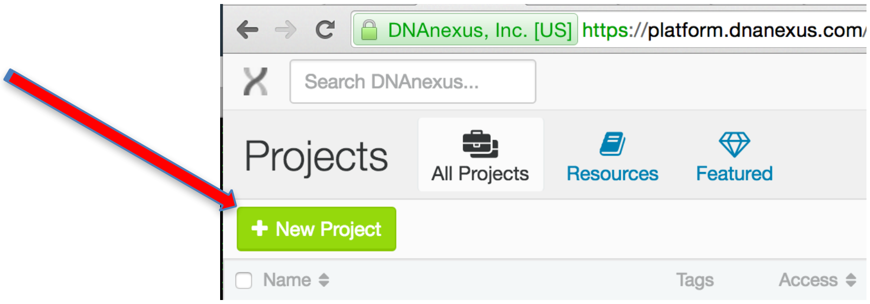
Summary of Steps: Here is a high-level summary of what you will learn to do in this exercise.

- Find the ENCODE Uniform Processing Pipeline project on DNAnexus.
- Copy the pipeline software and files from that project to a new project in your account.
- Complete the specification of inputs to the workflow.
- Run the pipeline workflow on the cloud.
- Monitor the run's progress.
- Visualize the output.

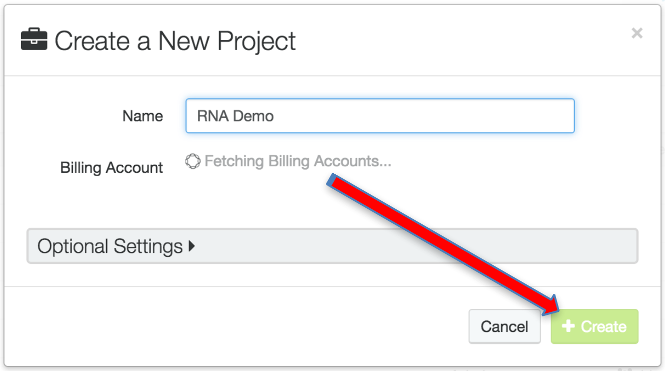
Step-by-step:

1) You will need to create an account on the DNAnexus website www.dnanexus.com. Log in to your DNAnexus account.

2) Once logged into your DNAnexus account, create a new project. Select "All Projects" and then click "New Project":



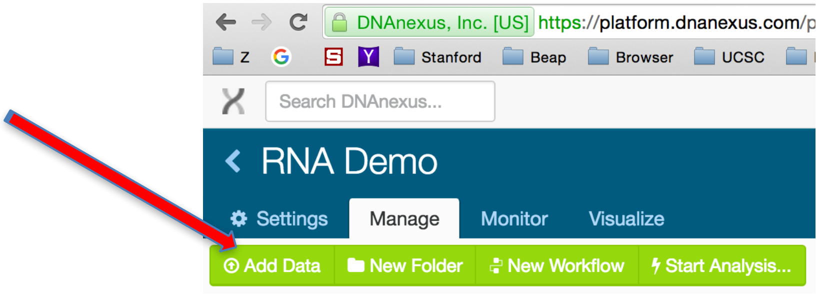
3) Give your project a new name and click “Create”.



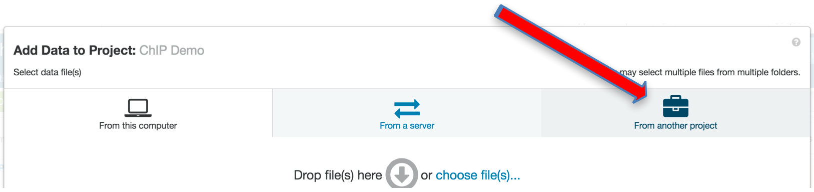
Name: RNA Demo

Billing Account: Fetching Billing Accounts...

4) Select “Add Data” …



5) … select “From another project” …

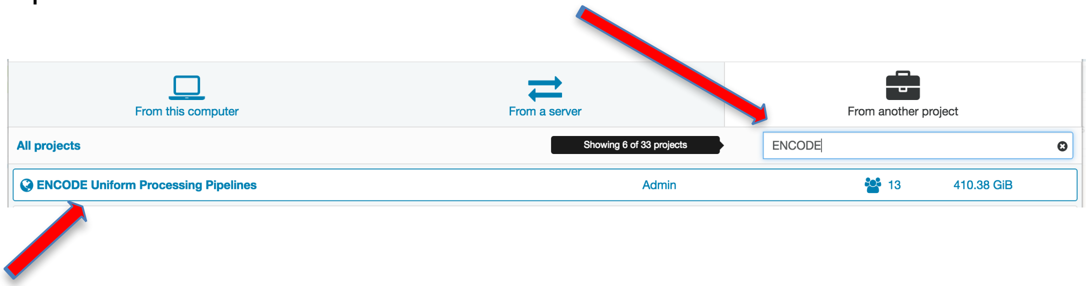


Add Data to Project: ChIP Demo

Select data file(s)

may select multiple files from multiple folders.

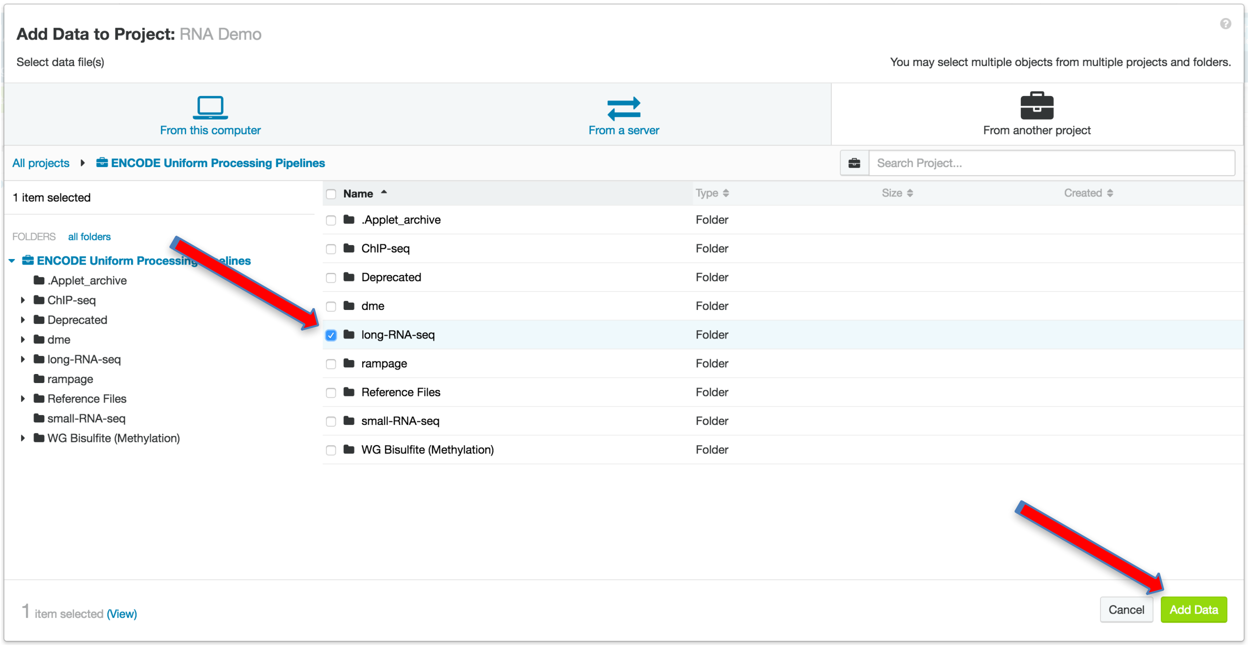
6) Type “ENCODE” in the search box and then select “ENCODE Uniform Processing Pipelines”



ENCODE

ENCODE Uniform Processing Pipelines	Admin	13	410.38 GiB

7) Click the box next to “long-RNA-seq” and select “Add Data”.



Add Data to Project: RNA Demo

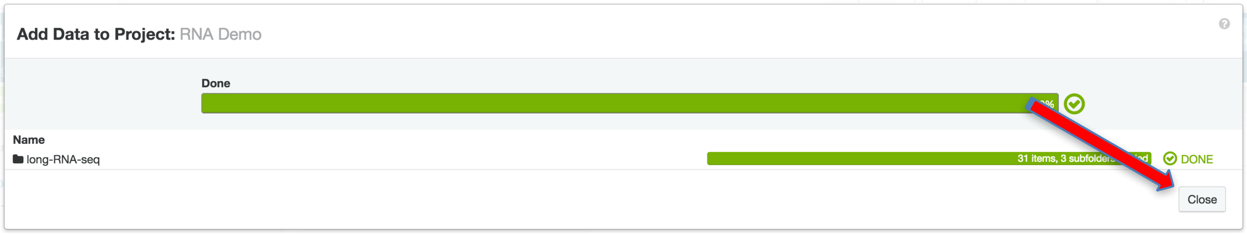
Select data file(s)

You may select multiple objects from multiple projects and folders.

	Name	Type	Size	Created
□	.Applet_archive	Folder		
□	ChIP-seq	Folder		
□	Deprecated	Folder		
□	dme	Folder		
☑	long-RNA-seq	Folder		
□	rampage	Folder		
□	Reference Files	Folder		
□	small-RNA-seq	Folder		
□	WG Bisulfite (Methylation)	Folder		

1 item selected (View)

8) When finished, the following pop-up window should appear. Click “Close”.



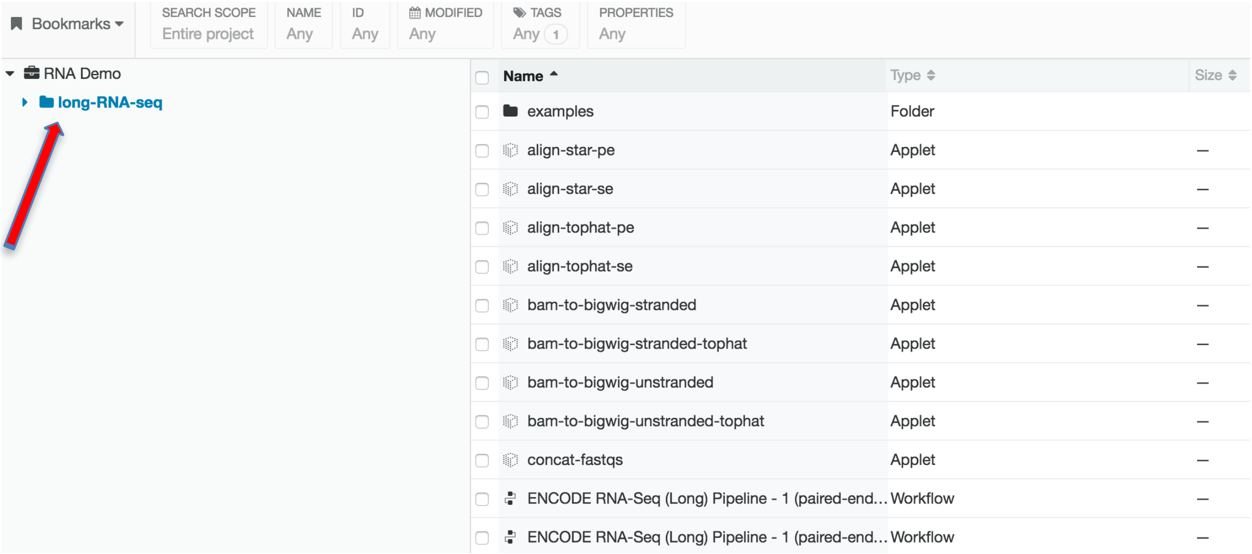
Add Data to Project: RNA Demo

Done

Name

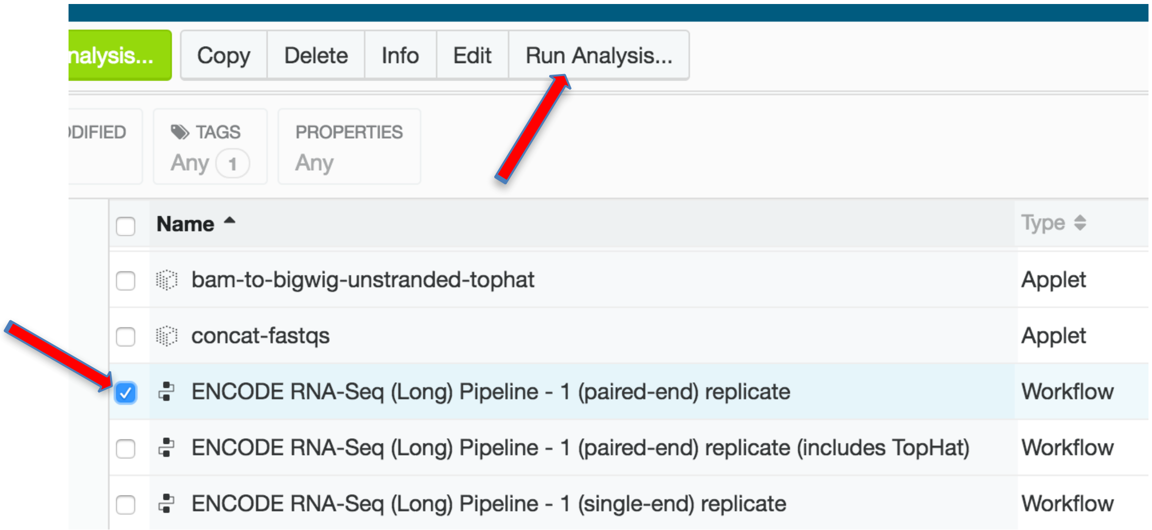
long-RNA-seq — 31 items, 3 subfolders ...ed — DONE

9) Click the long-RNA-seq text to open the folder. You should see the elements of the pipeline copied to your project.



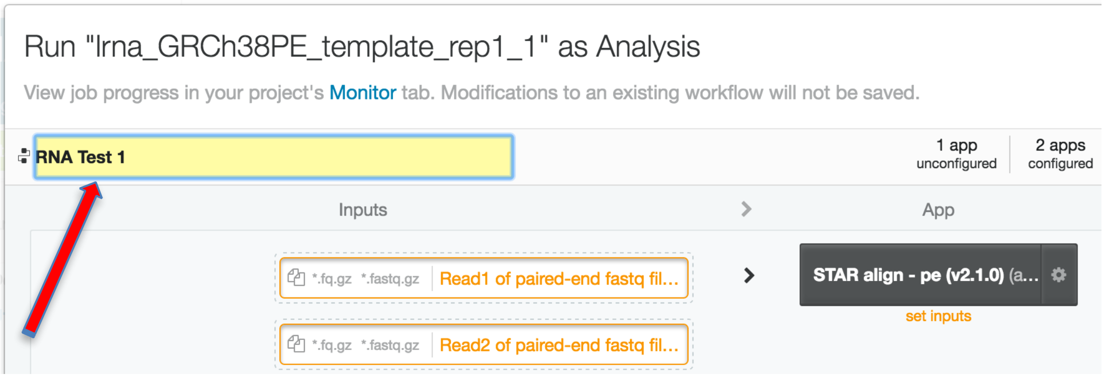
	Name	Type	Size
□	examples	Folder	
□	align-star-pe	Applet	—
□	align-star-se	Applet	—
□	align-tophat-pe	Applet	—
□	align-tophat-se	Applet	—
□	bam-to-bigwig-stranded	Applet	—
□	bam-to-bigwig-stranded-tophat	Applet	—
□	bam-to-bigwig-unstranded	Applet	—
□	bam-to-bigwig-unstranded-tophat	Applet	—
□	concat-fastqs	Applet	—
□	ENCODE RNA-Seq (Long) Pipeline - 1 (paired-end...	Workflow	—
□	ENCODE RNA-Seq (Long) Pipeline - 1 (paired-end...	Workflow	—

10) Select the workflow named “ENCODE RNA-seq (Long) Pipeline – 1 replicate (paired-end)”. (You may need to resize the “Name” column and scroll to distinguish among the several versions of the pipeline.) Upon selecting the workflow, press “Run Analysis…”.



□	Name	Type
□	bam-to-bigwig-unstranded-tophat	Applet
□	concat-fastqs	Applet
☑	ENCODE RNA-Seq (Long) Pipeline - 1 (paired-end) replicate	Workflow
□	ENCODE RNA-Seq (Long) Pipeline - 1 (paired-end) replicate (includes TopHat)	Workflow
□	ENCODE RNA-Seq (Long) Pipeline - 1 (single-end) replicate	Workflow

11) This window represents an “Analysis”, which is an instance of the long-RNA-seq workflow. Give the analysis an informative name, like “Total RNA for chr21 of human fetal stomach tissue”, or “RNA Test 1”.

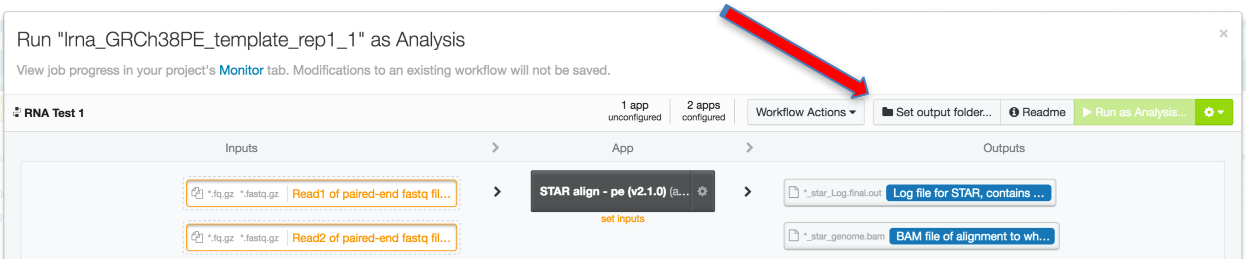

Run "lrna_GRCh38PE_template_rep1_1" as Analysis

View job progress in your project's Monitor tab. Modifications to an existing workflow will not be saved.

RNA Test 1

1 app unconfigured
2 apps configured
Inputs
App
*.fq.gz *.fastq.gz Read1 of paired-end fastq fil...
*.fq.gz *.fastq.gz Read2 of paired-end fastq fil...
STAR align - pe (v2.1.0) (a...
set inputs

12) Next, select a folder where your result files should be placed. Click on “Set output folder …”

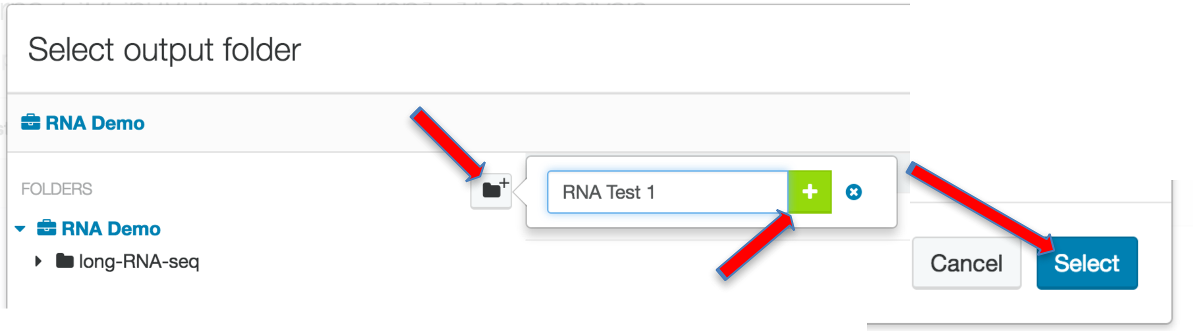

Run "lrna_GRCh38PE_template_rep1_1" as Analysis

View job progress in your project's Monitor tab. Modifications to an existing workflow will not be saved.

RNA Test 1

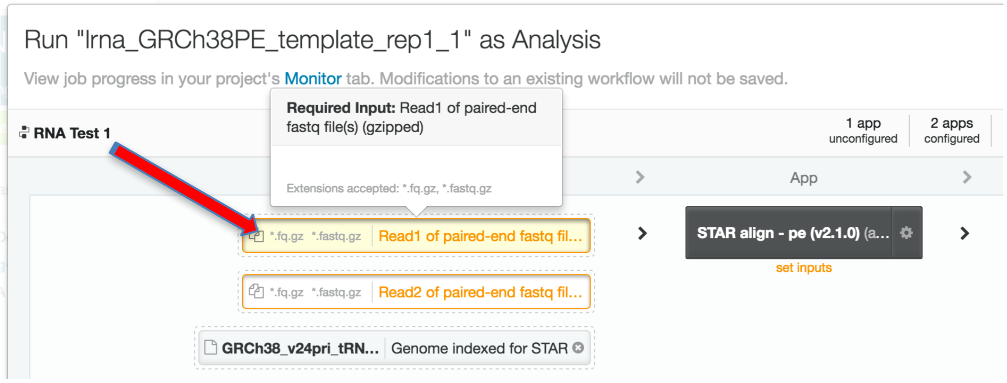
1 app unconfigured
2 apps configured
Workflow Actions
Set output folder...
Readme
Run as Analysis...
Inputs
App
Outputs
*.fq.gz *.fastq.gz Read1 of paired-end fastq fil...
*.fq.gz *.fastq.gz Read2 of paired-end fastq fil...
STAR align - pe (v2.1.0) (a...
set inputs
*_star_Log.final.out Log file for STAR, contains ...
*_star_genome.bam BAM file of alignment to wh...

13) Click on the new folder button to create a new folder and name it something catchy like "Chromosome 21 of ENCODE experiment ENCSR000AFI" or "RNA Test 1". Be sure to press the "+" sign and then "Select" at lower right.



Select output folder

14) Now it is time to add the sample data files to your analysis. Select the input box with orange text that says "Read1 of paired-end fast file(s) (gzipped)".



Run "lrna_GRCh38PE_template_rep1_1" as Analysis

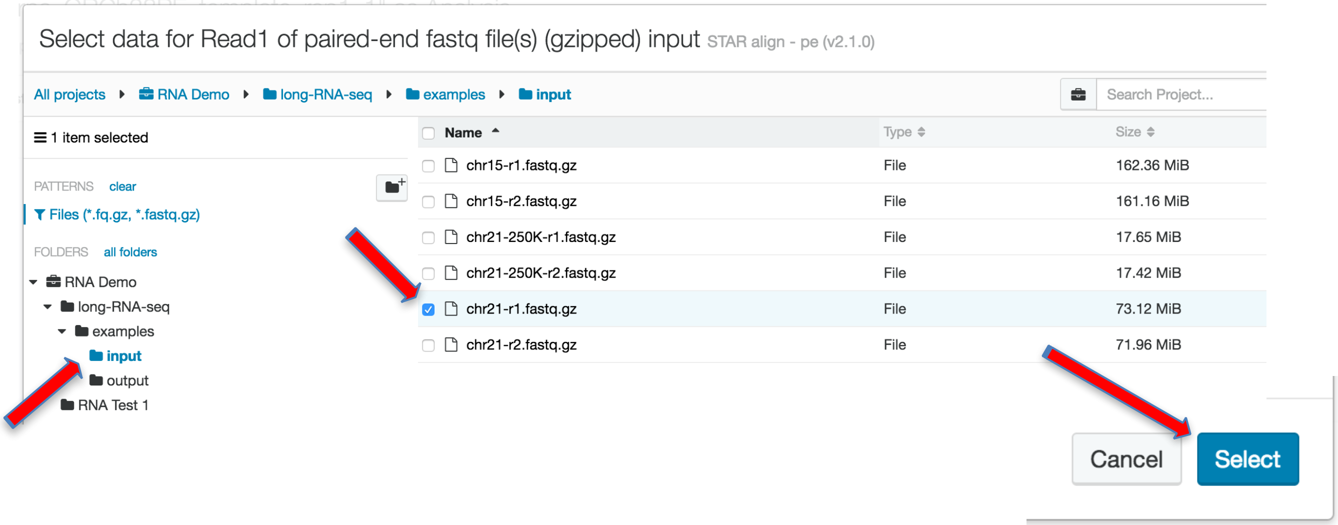
View job progress in your project's Monitor tab. Modifications to an existing workflow will not be saved.

Required Input: Read1 of paired-end fastq file(s) (gzipped)

Extensions accepted: *.fq.gz, *.fastq.gz

RNA Test 1
1 app unconfigured
2 apps configured
App
*.fq.gz *.fastq.gz Read1 of paired-end fastq fil...
*.fq.gz *.fastq.gz Read2 of paired-end fastq fil...
GRCh38_v24pri_tRN... Genome indexed for STAR
STAR align - pe (v2.1.0) (a...
set inputs

15) A new window opens where you will navigate to the input files. Expand the "long-RNA-seq" folder and within that "examples" and then "input". From here select the file named "chr21-r1.fastq.gz", then press "Select".



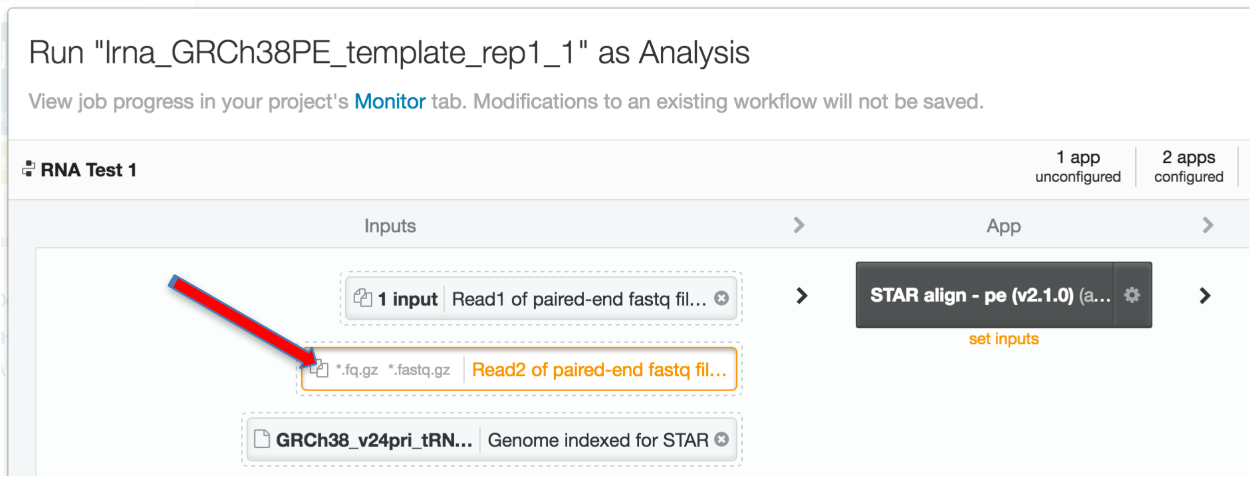
Select data for Read1 of paired-end fastq file(s) (gzipped) input STAR align - pe (v2.1.0)

All projects ▸ RNA Demo ▸ long-RNA-seq ▸ examples ▸ input

1 item selected

	Name	Type	Size
□	chr15-r1.fastq.gz	File	162.36 MiB
□	chr15-r2.fastq.gz	File	161.16 MiB
□	chr21-250K-r1.fastq.gz	File	17.65 MiB
□	chr21-250K-r2.fastq.gz	File	17.42 MiB
☑	chr21-r1.fastq.gz	File	73.12 MiB
□	chr21-r2.fastq.gz	File	71.96 MiB

16) Now repeat for “Read2”, choosing “chr21-r2.fastq.gz”.



Run "lrna_GRCh38PE_template_rep1_1" as Analysis

View job progress in your project's Monitor tab. Modifications to an existing workflow will not be saved.

RNA Test 1

1 app unconfigured | 2 apps configured

Inputs | App

1 input | Read1 of paired-end fastq fil...

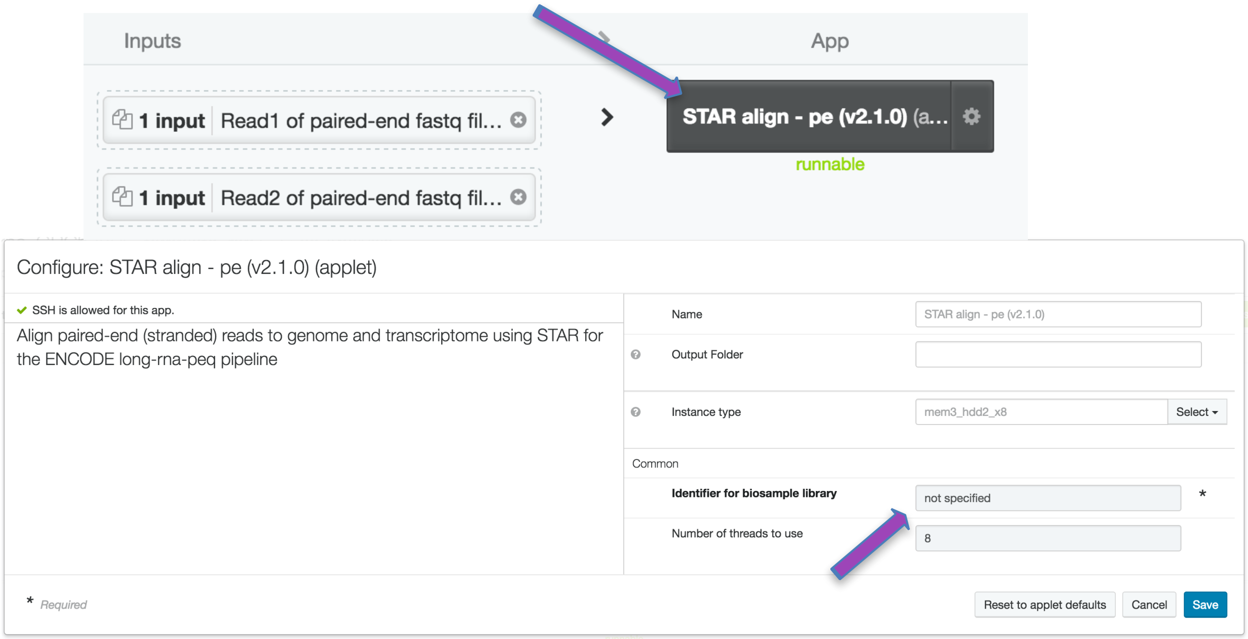
*.fq.gz *.fastq.gz | Read2 of paired-end fastq fil...

GRCh38_v24pri_tRN... | Genome indexed for STAR

STAR align - pe (v2.1.0) (a...

set inputs

17) Notice that there is no longer any orange text declaring required inputs. Nevertheless, there are *optional* inputs that one may declare. Select the box labeled “STAR align - pe (v2.1.0). Here you may enter an “Identifier for biosample library”, which will wind up being embedded in the result bam file’s header. There are other options for running this alignment step on DNAnexus. It is not recommended that you alter these.



Inputs | App

1 input | Read1 of paired-end fastq fil...

1 input | Read2 of paired-end fastq fil...

STAR align - pe (v2.1.0) (a...

runnable

Configure: STAR align - pe (v2.1.0) (applet)

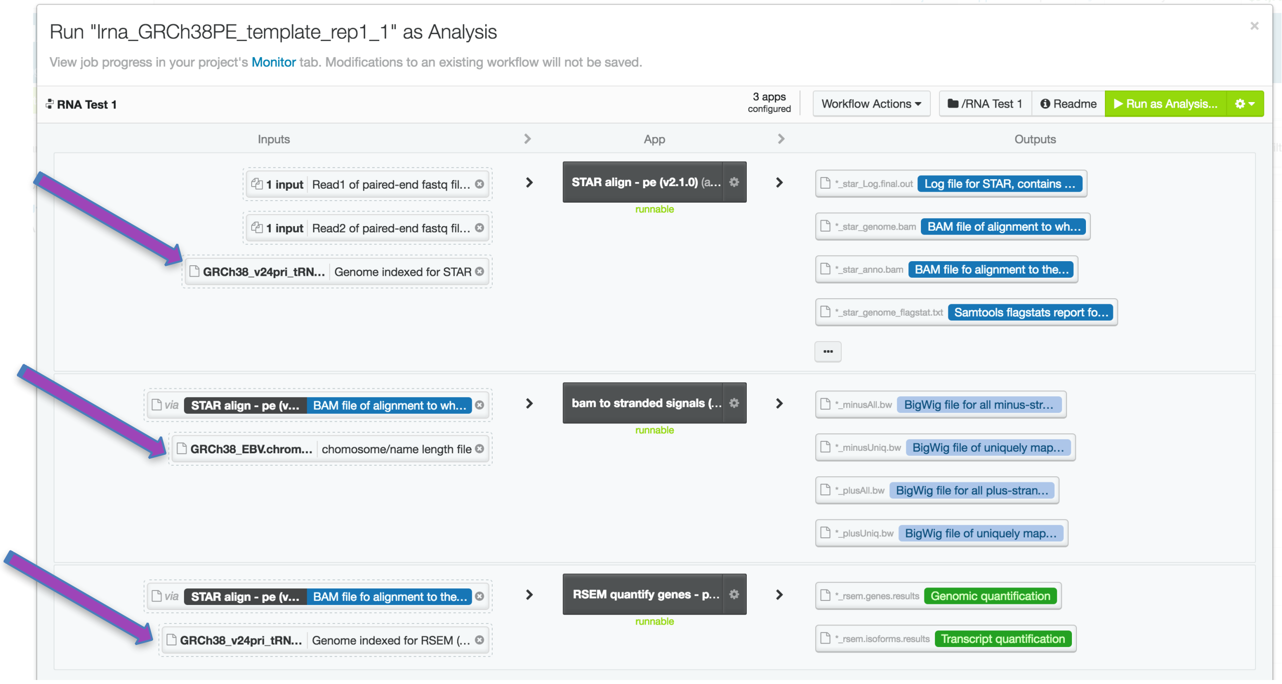
✓ SSH is allowed for this app.

Align paired-end (stranded) reads to genome and transcriptome using STAR for the ENCODE long-rna-peq pipeline

Name	STAR align - pe (v2.1.0)
Output Folder	
Instance type	mem3_hdd2_x8 Select ▾
Common	
Identifier for biosample library	not specified *
Number of threads to use	8

* Required

18) This analysis is already set to run on GRCh38. It may also be run on the hg19 assembly. To do this would require replacing three reference input files with the corresponding hg19 files, which are located in the same “ENCODE Uniform Processing Pipelines” project, from which you copied this pipeline.

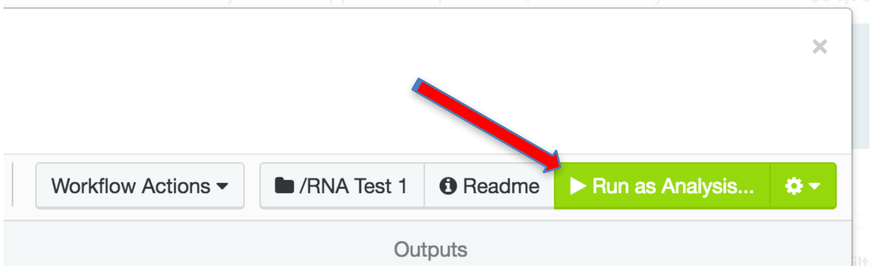

Run "lrna_GRCh38PE_template_rep1_1" as Analysis

View job progress in your project's Monitor tab. Modifications to an existing workflow will not be saved.

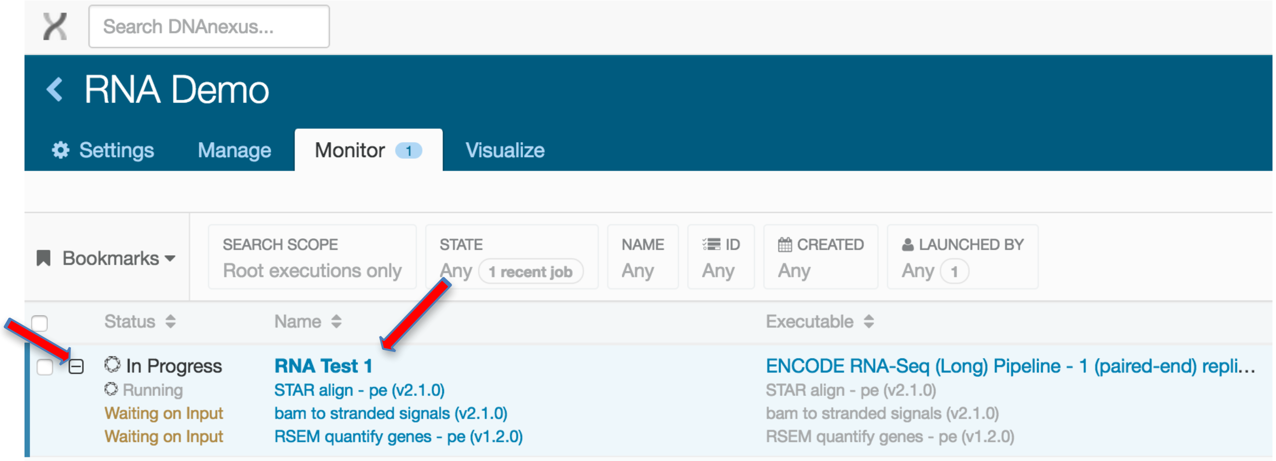
RNA Test 1

Inputs	App	Outputs
1 input Read1 of paired-end fastq fil... 1 input Read2 of paired-end fastq fil... GRCh38_v24pri_tRN... Genome indexed for STAR	STAR align - pe (v2.1.0) (a... runnable	*_star_Log.final.out Log file for STAR, contains ... *_star_genome.bam BAM file of alignment to wh... *_star_anno.bam BAM file fo alignment to the... *_star_genome_flagstat.txt Samtools flagstats report fo... ...
via STAR align - pe (v... BAM file of alignment to wh... GRCh38_EBV.chrom... chomosome/name length file	bam to stranded signals (... runnable	*_minusAll.bw BigWig file for all minus-str... *_minusUniq.bw BigWig file of uniquely map... *_plusAll.bw BigWig file for all plus-stran... *_plusUniq.bw BigWig file of uniquely map...
via STAR align - pe (v... BAM file fo alignment to the... GRCh38_v24pri_tRN... Genome indexed for RSEM (...	RSEM quantify genes - p... runnable	*_rsem.genes.results Genomic quantification *_rsem.isoforms.results Transcript quantification

19) With all the required inputs provided, it is time to run your analysis, so click “Run as Analysis”.



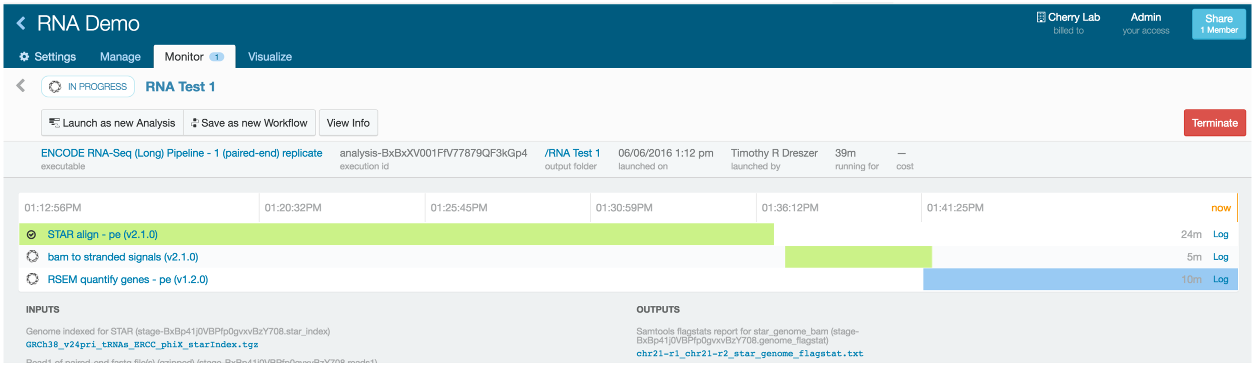
20) Starting the analysis will bring up the “Monitor” tab which will display the details of the pipeline steps as they run. Click on the “+” box to see the analysis subjobs. If necessary, the “Terminate” button can be used to cancel the analysis.



SEARCH SCOPE Root executions only | STATE Any 1 recent job | NAME Any | ID Any | CREATED Any | LAUNCHED BY Any 1

Status	Name	Executable
In Progress	RNA Test 1	ENCODE RNA-Seq (Long) Pipeline - 1 (paired-end) repli...
Running	STAR align - pe (v2.1.0)	STAR align - pe (v2.1.0)
Waiting on Input	bam to stranded signals (v2.1.0)	bam to stranded signals (v2.1.0)
Waiting on Input	RSEM quantify genes - pe (v1.2.0)	RSEM quantify genes - pe (v1.2.0)

21) Click on the analysis name to watch the progress of each step.



ENCODE RNA-Seq (Long) Pipeline - 1 (paired-end) replicate	analysis-BxBxXV001FfV77879QF3kGp4	/RNA Test 1	06/06/2016 1:12 pm	Timothy R Dreszer	39m	—
executable	execution id	output folder	launched on	launched by	running for	cost

01:12:56PM 01:20:32PM 01:25:45PM 01:30:59PM 01:36:12PM 01:41:25PM now

STAR align - pe (v2.1.0)	24m	Log
bam to stranded signals (v2.1.0)	5m	Log
RSEM quantify genes - pe (v1.2.0)	10m	Log

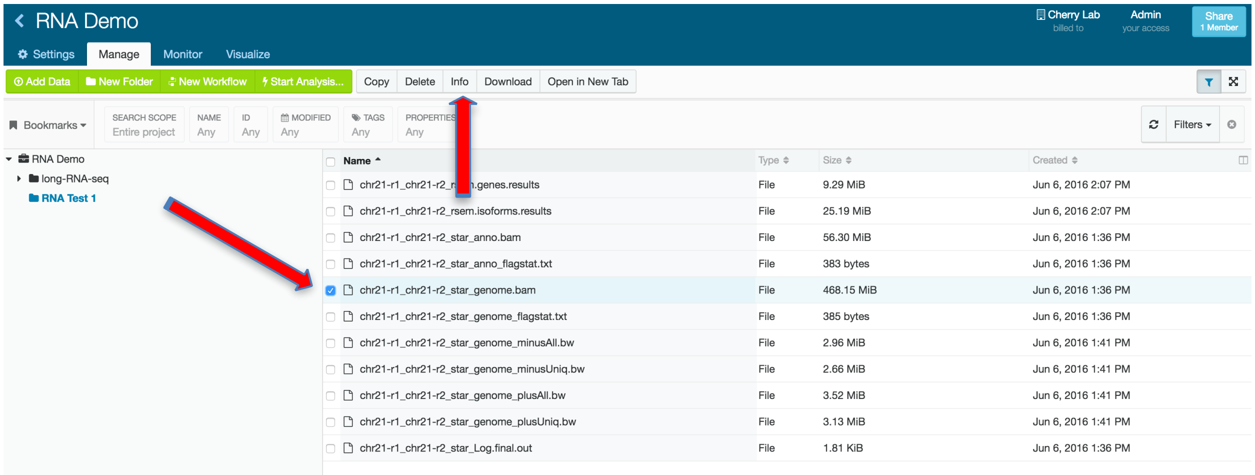
INPUTS

Genome indexed for STAR (stage-BxBp41j0VBPfp0gvxvBzY708.star_index)
`GRCh38_v24pri_tRNAs_ERCC_phiX_starIndex.tgz`

OUTPUTS

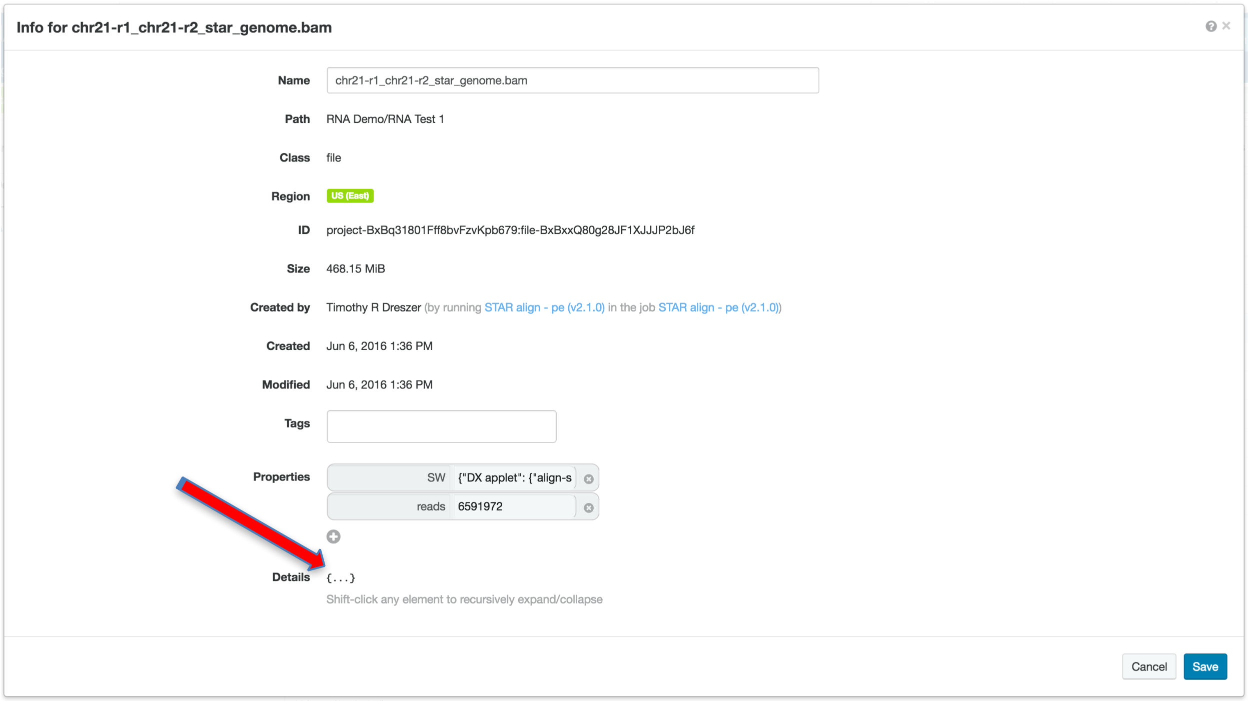
Samtools flagstats report for star_genome_bam (stage-BxBp41j0VBPfp0gvxvBzY708.genome_flagstat)
`chr21-r1_chr21-r2_star_genome_flagstat.txt`

22) Within the output folder you specified above, all result files will accumulate as the steps of the pipeline complete. Many of the files will have additional information. For example, select the file whose name ends with "_star_genome.bam", then choose "Info".



Name	Type	Size	Created
☐ chr21-r1_chr21-r2_r[illegible]n.genes.results	File	9.29 MiB	Jun 6, 2016 2:07 PM
☐ chr21-r1_chr21-r2_rsem.isoforms.results	File	25.19 MiB	Jun 6, 2016 2:07 PM
☐ chr21-r1_chr21-r2_star_anno.bam	File	56.30 MiB	Jun 6, 2016 1:36 PM
☐ chr21-r1_chr21-r2_star_anno_flagstat.txt	File	383 bytes	Jun 6, 2016 1:36 PM
☑ chr21-r1_chr21-r2_star_genome.bam	File	468.15 MiB	Jun 6, 2016 1:36 PM
☐ chr21-r1_chr21-r2_star_genome_flagstat.txt	File	385 bytes	Jun 6, 2016 1:36 PM
☐ chr21-r1_chr21-r2_star_genome_minusAll.bw	File	2.96 MiB	Jun 6, 2016 1:41 PM
☐ chr21-r1_chr21-r2_star_genome_minusUniq.bw	File	2.66 MiB	Jun 6, 2016 1:41 PM
☐ chr21-r1_chr21-r2_star_genome_plusAll.bw	File	3.52 MiB	Jun 6, 2016 1:41 PM
☐ chr21-r1_chr21-r2_star_genome_plusUniq.bw	File	3.13 MiB	Jun 6, 2016 1:41 PM
☐ chr21-r1_chr21-r2_star_Log.final.out	File	1.81 KiB	Jun 6, 2016 1:36 PM

23) This alignment file reports 6,591,972 reads but there is even more detail to be found by clicking on the "{…}".

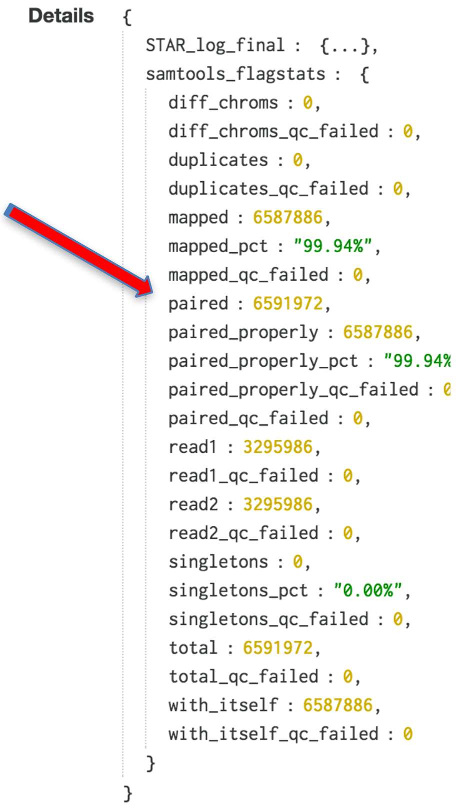


Info for chr21-r1_chr21-r2_star_genome.bam

Name	chr21-r1_chr21-r2_star_genome.bam
Path	RNA Demo/RNA Test 1
Class	file
Region	US (East)
ID	project-BxBq31801Ff8bvFzvKpb679:file-BxBxxQ80g28JF1XJJJP2bJ6f
Size	468.15 MiB
Created by	Timothy R Dreszer (by running STAR align - pe (v2.1.0) in the job STAR align - pe (v2.1.0))
Created	Jun 6, 2016 1:36 PM
Modified	Jun 6, 2016 1:36 PM
Tags	
Properties	SW {"DX applet": {"align-s reads 6591972
Details	{...}

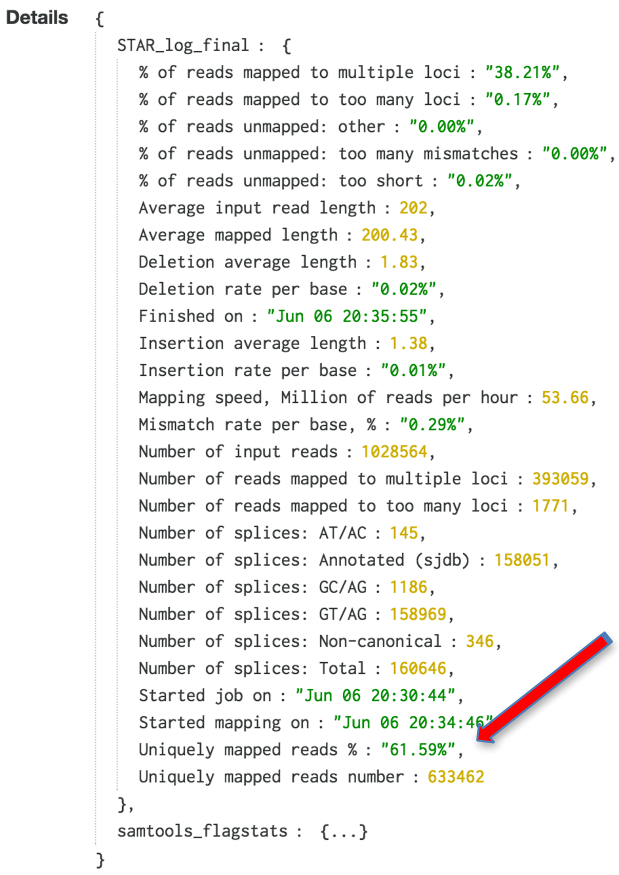
Shift-click any element to recursively expand/collapse

24) By expanding on the “samtools_flagstats {…}”, we learn that 6,591,972 is actually read pairs, and by expanding “STAR_final_log {…}”, we can learn that 61.59% were uniquely mapped.



Details

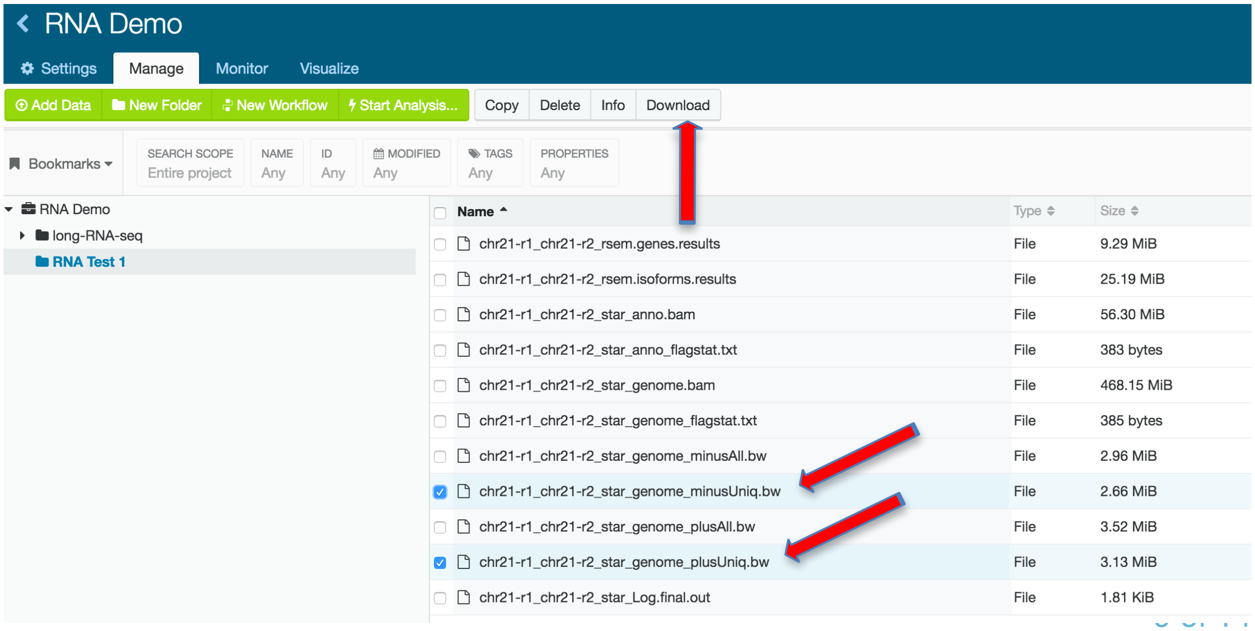
```
{
  STAR_log_final :  {...},
  samtools_flagstats :  {
    diff_chroms : 0,
    diff_chroms_qc_failed : 0,
    duplicates : 0,
    duplicates_qc_failed : 0,
    mapped : 6587886,
    mapped_pct : "99.94%",
    mapped_qc_failed : 0,
    paired : 6591972,
    paired_properly : 6587886,
    paired_properly_pct : "99.94%
    paired_properly_qc_failed : 0
    paired_qc_failed : 0,
    read1 : 3295986,
    read1_qc_failed : 0,
    read2 : 3295986,
    read2_qc_failed : 0,
    singletons : 0,
    singletons_pct : "0.00%",
    singletons_qc_failed : 0,
    total : 6591972,
    total_qc_failed : 0,
    with_itself : 6587886,
    with_itself_qc_failed : 0
  }
}
```



Details

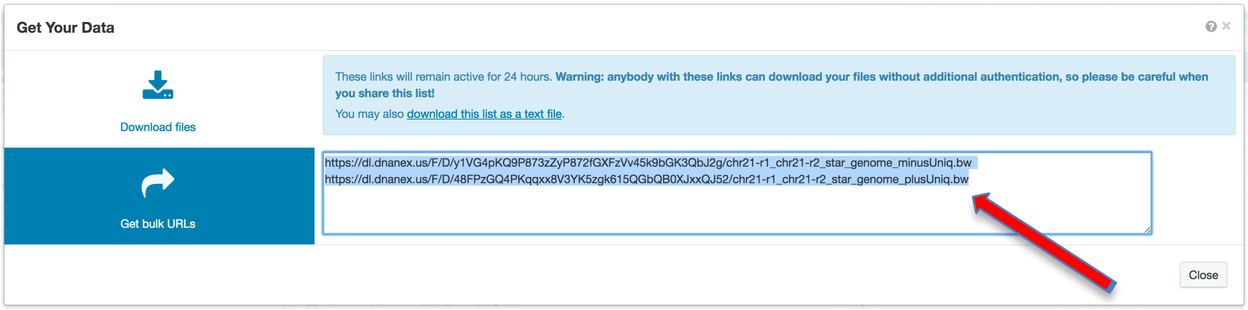
```
{
  STAR_log_final :  {
    % of reads mapped to multiple loci : "38.21%",
    % of reads mapped to too many loci : "0.17%",
    % of reads unmapped: other : "0.00%",
    % of reads unmapped: too many mismatches : "0.00%",
    % of reads unmapped: too short : "0.02%",
    Average input read length : 202,
    Average mapped length : 200.43,
    Deletion average length : 1.83,
    Deletion rate per base : "0.02%",
    Finished on : "Jun 06 20:35:55",
    Insertion average length : 1.38,
    Insertion rate per base : "0.01%",
    Mapping speed, Million of reads per hour : 53.66,
    Mismatch rate per base, % : "0.29%",
    Number of input reads : 1028564,
    Number of reads mapped to multiple loci : 393059,
    Number of reads mapped to too many loci : 1771,
    Number of splices: AT/AC : 145,
    Number of splices: Annotated (sjdb) : 158051,
    Number of splices: GC/AG : 1186,
    Number of splices: GT/AG : 158969,
    Number of splices: Non-canonical : 346,
    Number of splices: Total : 160646,
    Started job on : "Jun 06 20:30:44",
    Started mapping on : "Jun 06 20:34:46",
    Uniquely mapped reads % : "61.59%",
    Uniquely mapped reads number : 633462
  },
  samtools_flagstats :  {...}
}
```

25) To visualize the signal results as custom tracks at the UCSC Genome Browser, select the the 2 bigwig files ending in “_minusUniq.bw” and “_plusUniq.bw”. These two files are the signal produced from only uniquely mapped reads for the minus and plus DNA strands respectively. Select “Download.”



	Name	Type	Size
□	chr21-r1_chr21-r2_rsem.genes.results	File	9.29 MiB
□	chr21-r1_chr21-r2_rsem.isoforms.results	File	25.19 MiB
□	chr21-r1_chr21-r2_star_anno.bam	File	56.30 MiB
□	chr21-r1_chr21-r2_star_anno_flagstat.txt	File	383 bytes
□	chr21-r1_chr21-r2_star_genome.bam	File	468.15 MiB
□	chr21-r1_chr21-r2_star_genome_flagstat.txt	File	385 bytes
□	chr21-r1_chr21-r2_star_genome_minusAll.bw	File	2.96 MiB
☑	chr21-r1_chr21-r2_star_genome_minusUniq.bw	File	2.66 MiB
□	chr21-r1_chr21-r2_star_genome_plusAll.bw	File	3.52 MiB
☑	chr21-r1_chr21-r2_star_genome_plusUniq.bw	File	3.13 MiB
□	chr21-r1_chr21-r2_star_Log.final.out	File	1.81 KiB

26) A new window will pop up. Select “Get bulk URLs” and copy the two URLs. These URL’s will link to your output files and will remain active for 24 hours.

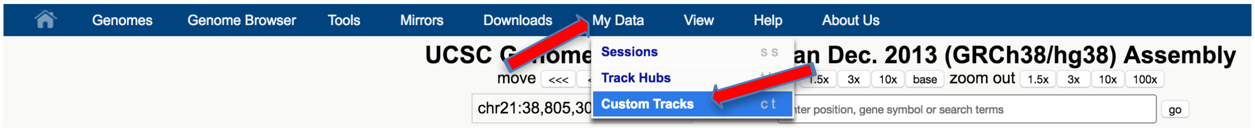


These links will remain active for 24 hours. **Warning: anybody with these links can download your files without additional authentication, so please be careful when you share this list!**

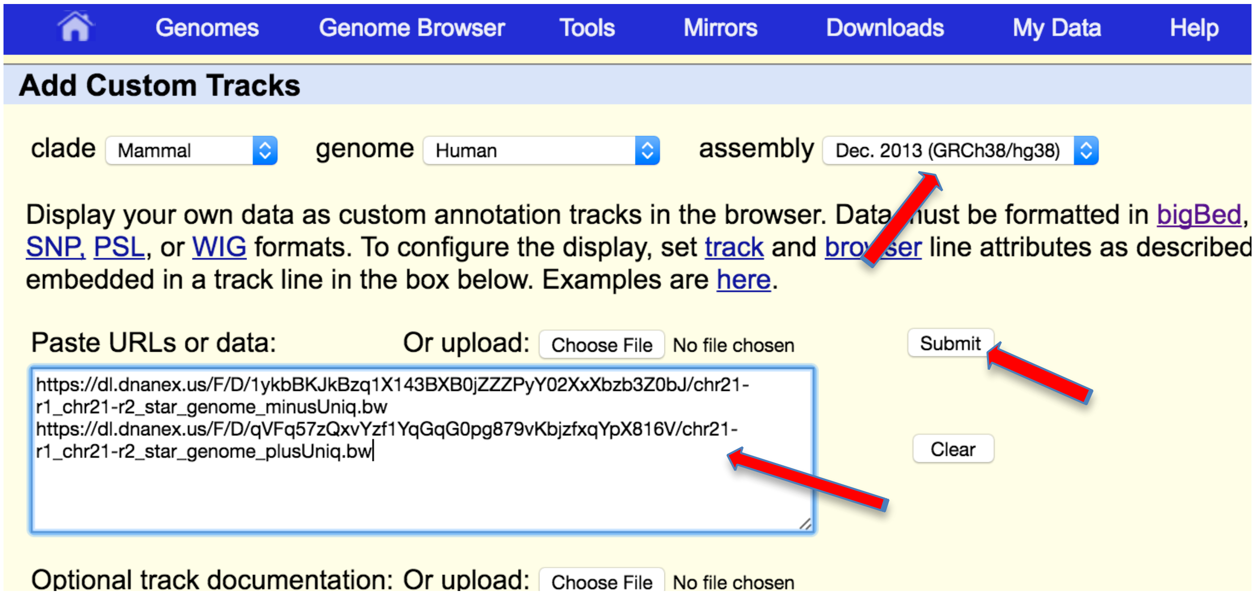
You may also download this list as a text file.

```
https://dl.dnanex.us/F/D/y1VG4pKQ9P873zZyP872fGXFzVv45k9bGK3QbJ2g/chr21-r1_chr21-r2_star_genome_minusUniq.bw
https://dl.dnanex.us/F/D/48FPzGQ4PKqqxx8V3YK5zgk615QGbQB0XJxxQJ52/chr21-r1_chr21-r2_star_genome_plusUniq.bw
```

27) In a new web browser window or tab, go to http://genome.ucsc.edu/ and select “My Data” from the top options bar, then select “Custom Tracks”.



28) Paste the URLs you copied above into the first text window. Be sure the reference genome is correct for your results (human GRCh38/hg38 for this demo). ***Tip: The UCSC Genome Browser is sensitive to white-space at the end of URL’s. If there are spaces after the URL’s you’ve pasted, delete them and make sure each URL is on its own line.*** Now press “Submit”.



Add Custom Tracks

clade Mammal genome Human assembly Dec. 2013 (GRCh38/hg38)

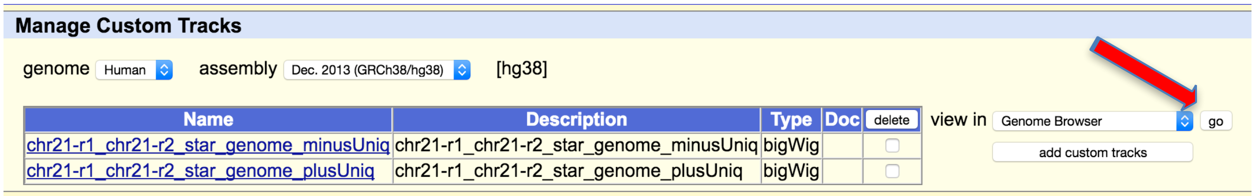
Display your own data as custom annotation tracks in the browser. Data ust be formatted in bigBed, SNP, PSL, or WIG formats. To configure the display, set track and bro ser line attributes as described embedded in a track line in the box below. Examples are here.

Paste URLs or data: Or upload:

```
https://dl.dnanex.us/F/D/1ykbBKJkBzq1X143BXB0jZZZPyY02XxXbzb3Z0bJ/chr21-
r1_chr21-r2_star_genome_minusUniq.bw
https://dl.dnanex.us/F/D/qVFq57zQxvYzf1YqGqG0pg879vKbjzfxqYpX816V/chr21-
r1_chr21-r2_star_genome_plusUniq.bw
```

Optional track documentation: Or upload:

29) You have one more chance to verify that you have selected GRCh38, before pressing “go”.

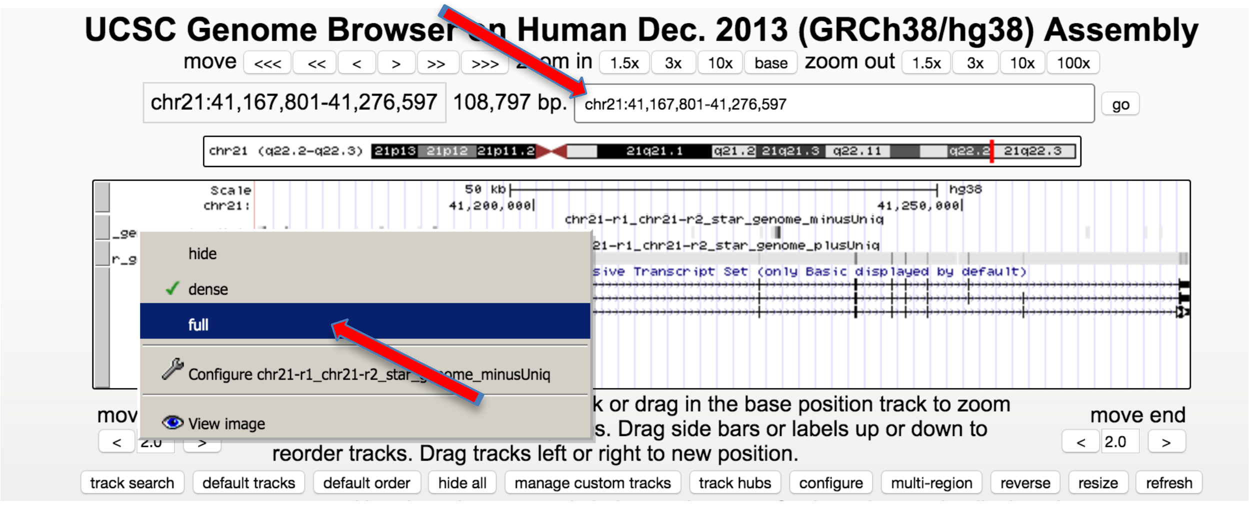


Manage Custom Tracks

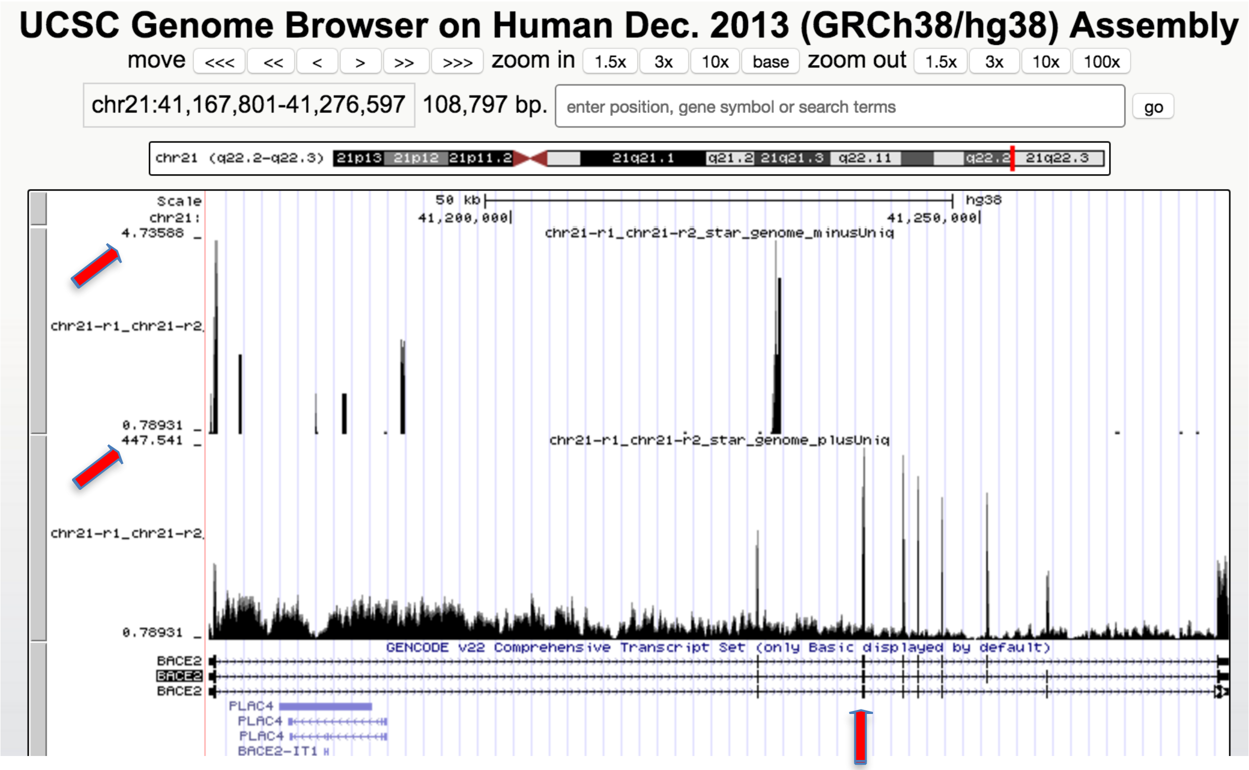
genome Human assembly Dec. 2013 (GRCh38/hg38) [hg38]

Name	Description	Type	Doc	delete
chr21-r1_chr21-r2_star_genome_minusUniq	chr21-r1_chr21-r2_star_genome_minusUniq	bigWig		□
chr21-r1_chr21-r2_star_genome_plusUniq	chr21-r1_chr21-r2_star_genome_plusUniq	bigWig		□

30) Your two custom tracks will be displayed at the top of the browser image. Because the raw data were subsampled to only chromosome 21, there should be no significant result anywhere else. Set the browser's position to chr21:41,167,801-41,276,597. This is the location of the BACE2 gene. To see more clearly your results, change both of the custom tracks to "full" (right-click on the track in the image).


UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
chr21:41,167,801-41,276,597 108,797 bp. chr21:41,167,801-41,276,597 go
chr21 (q22.2-q22.3) 21p13 21p12 21p11.2 21q21.1 q21.2 21q21.3 q22.11 q22.2 21q22.3
Scale chr21: 50 kb hg38 41,200,000| 41,250,000|
chr21-r1_chr21-r2_star_genome_minusUniq
21-r1_chr21-r2_star_genome_plusUniq
sive Transcript Set (only Basic displayed by default)
hide
✓ dense
full
Configure chr21-r1_chr21-r2_star_genome_minusUniq
View image
k or drag in the base position track to zoom
s. Drag side bars or labels up or down to
reorder tracks. Drag tracks left or right to new position.
move end < 2.0 >
track search default tracks default order hide all manage custom tracks track hubs configure multi-region reverse resize refresh

31) The UCSC Browser image should clearly show signal spiking at the exons of BACE2 for only one of your two custom tracks. Notice that these signal tracks are autoscaling, so while the plus signal peaks at 447 RPM (reads per million mapped reads), the minus signal peaks at less than 5 RPM in this location. Try other genes located on chromosome 21. For example, HLCS, SOD1, ETS2, or AIRE.

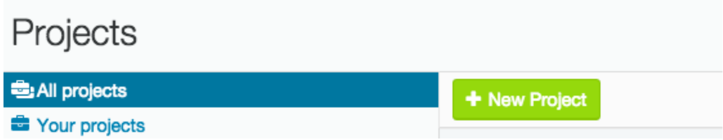

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
chr21:41,167,801-41,276,597 108,797 bp. enter position, gene symbol or search terms go
chr21 (q22.2-q22.3) 21p13 21p12 21p11.2 21q21.1 q21.2 21q21.3 q22.11 q22.2 21q22.3
Scale chr21: 50 kb hg38 41,200,000| 41,250,000|
4.73588
chr21-r1_chr21-r2_star_genome_minusUniq
chr21-r1_chr21-r2
0.78931
447.541
chr21-r1_chr21-r2_star_genome_plusUniq
chr21-r1_chr21-r2
0.78931
GENCODE v22 Comprehensive Transcript Set (only Basic displayed by default)
BACE2
BACE2
BACE2
PLAC4
PLAC4
PLAC4
BACE2-IT1

Congratulations! You have replicated an ENCODE analysis starting with primary data. You can repeat this process on your own data, and be assured that your results will be directly comparable to all the experiments the ENCODE DCC has analyzed.

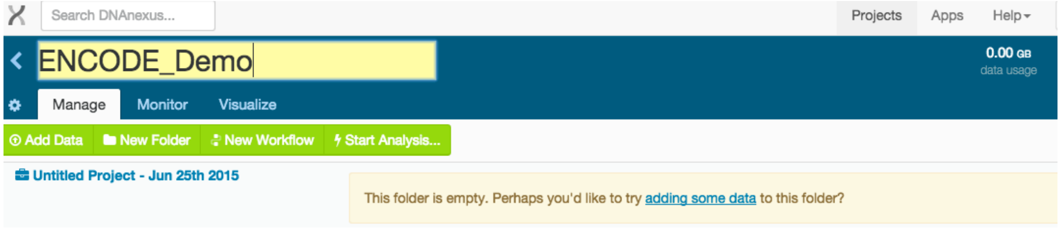
Other DNAnexus Tools:

To load data once you are in your own project

1) Start a "New Project" or find your own project in the DNAnexus homepage.



2) If new, name project in the upper left corner.

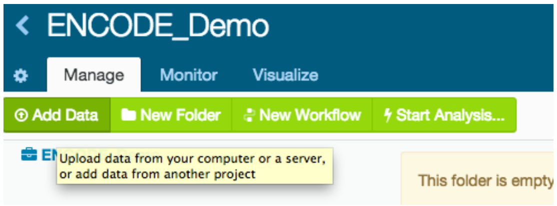

ENCODE_Demo

0.00 GB data usage

Untitled Project - Jun 25th 2015

This folder is empty. Perhaps you'd like to try adding some data to this folder?

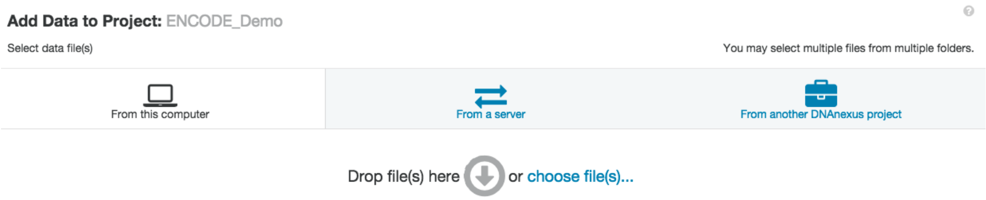
3) Select "Add Data" to select the files you want to use for analysis to your project.


ENCODE_Demo

Upload data from your computer or a server, or add data from another project

This folder is empty

4) When the "Add Data to Project" window pops up, select "From another DNAnexus project."

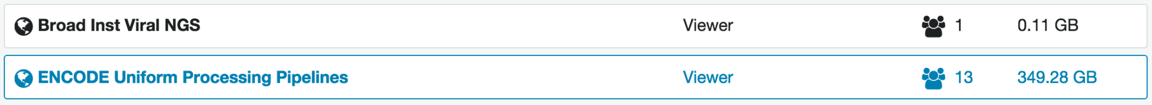

Add Data to Project: ENCODE_Demo

Select data file(s)

You may select multiple files from multiple folders.

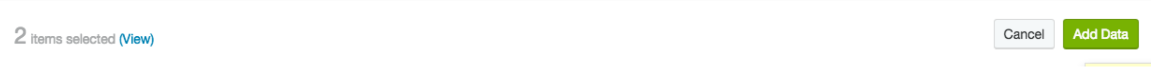
Drop file(s) here or choose file(s)...

5) Scroll down and select "ENCODE Universal Processing Pipeline" project to access the data.



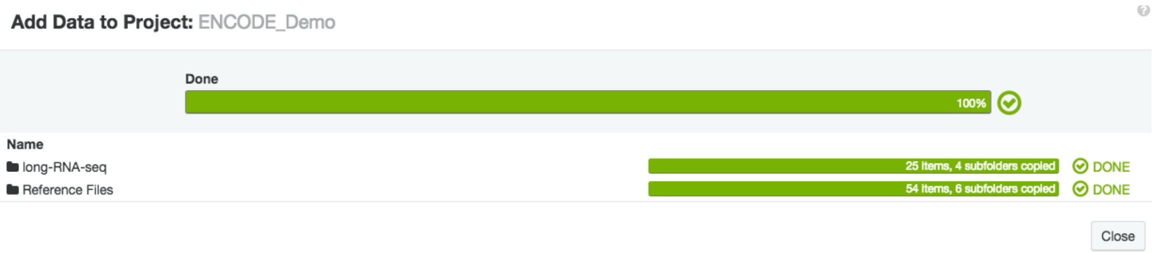
Broad Inst Viral NGS	Viewer	1	0.11 GB
ENCODE Uniform Processing Pipelines	Viewer	13	349.28 GB

6) Choose "Add Data" to select these files.



2 items selected (View)

7) When these files are uploaded, the following window will pop up.

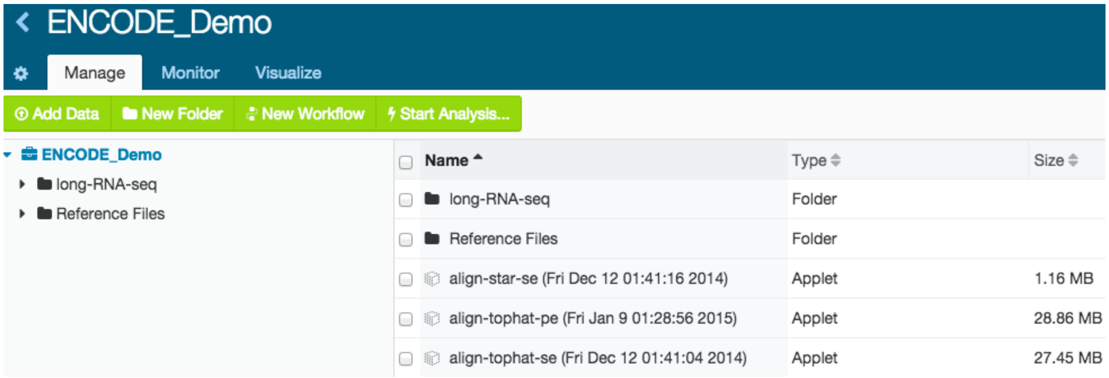


Add Data to Project: ENCODE_Demo

Done 100%

Name		
long-RNA-seq	25 items, 4 subfolders copied	DONE
Reference Files	54 items, 6 subfolders copied	DONE

8) These files and associated applets will now appear in the Manage tab of your browser.

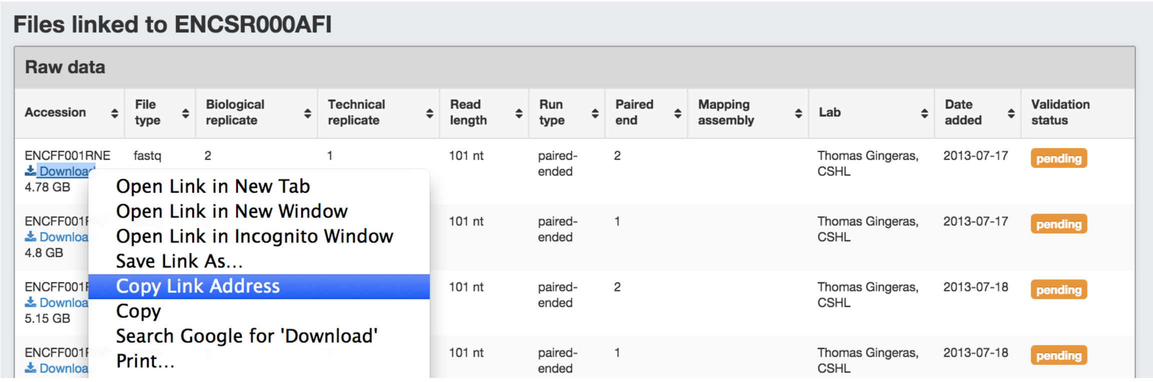


ENCODE_Demo

Name	Type	Size
long-RNA-seq	Folder	
Reference Files	Folder	
align-star-se (Fri Dec 12 01:41:16 2014)	Applet	1.16 MB
align-tophat-pe (Fri Jan 9 01:28:56 2015)	Applet	28.86 MB
align-tophat-se (Fri Dec 12 01:41:04 2014)	Applet	27.45 MB

To import a fastq file directly from the ENCODE portal to DNAnexus

1) Go to the ENCODE portal (encodeproject.org) and find the fastq file you are interested in using. Right click on this file and select "Copy Link Address."

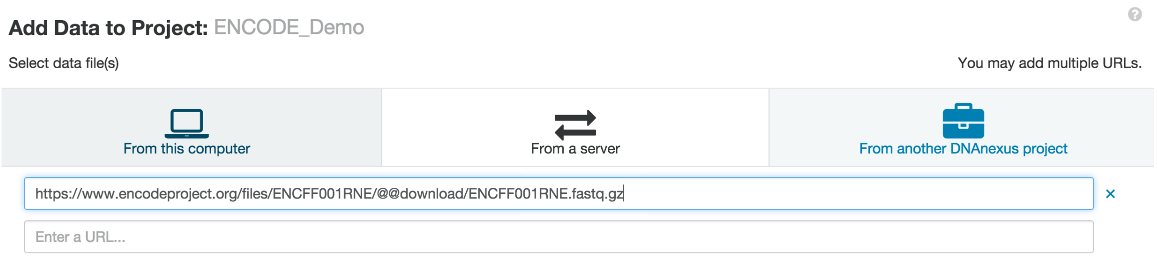


Files linked to ENCSR000AFI

Raw data

Accession	File type	Biological replicate	Technical replicate	Read length	Run type	Paired end	Mapping assembly	Lab	Date added	Validation status
ENCFF001RNE Downloa 4.78 GB	fastq	2	1	101 nt	paired-ended	2		Thomas Gingeras, CSHL	2013-07-17	pending
ENCFF001F Downloa 4.8 GB				101 nt	paired-ended	1		Thomas Gingeras, CSHL	2013-07-17	pending
ENCFF001F Downloa 5.15 GB				101 nt	paired-ended	2		Thomas Gingeras, CSHL	2013-07-18	pending
ENCFF001F Downloa				101 nt	paired-ended	1		Thomas Gingeras, CSHL	2013-07-18	pending

2) In the manage tab, under "Add Data" select the "From a Server" option and paste the URL into the box. Select "Add Data" and the file will upload.

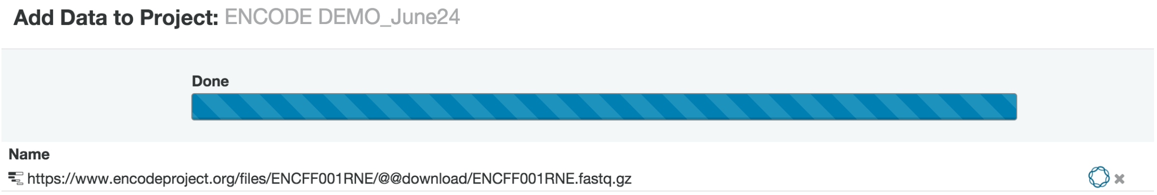


Add Data to Project: ENCODE_Demo

Select data file(s)

You may add multiple URLs.

https://www.encodeproject.org/files/ENCFF001RNE/@@download/ENCFF001RNE.fastq.gz



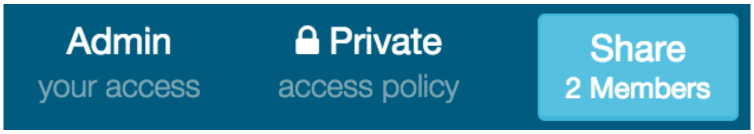
Add Data to Project: ENCODE DEMO_June24

Done

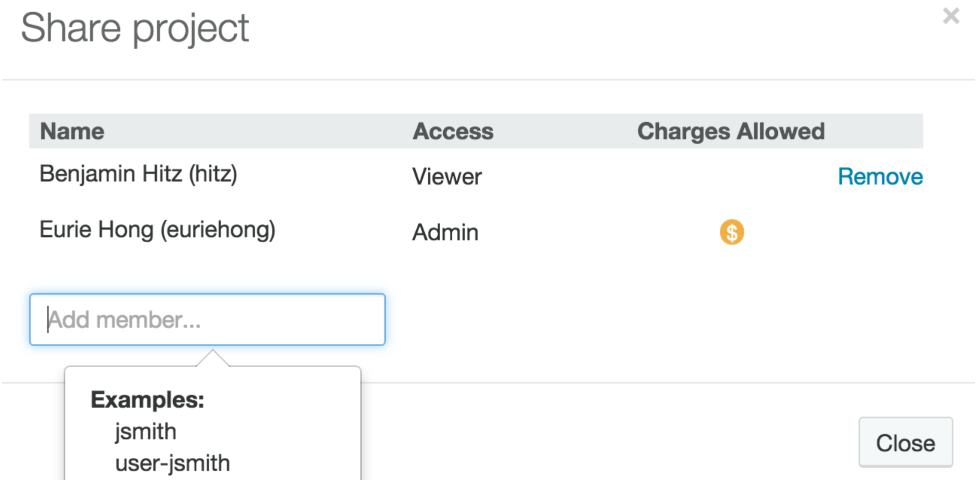
Name
https://www.encodeproject.org/files/ENCFF001RNE/@@download/ENCFF001RNE.fastq.gz

To share project with another user

1) In order to share your project, select the blue "Share" button at the upper right corner of the browser page.



2) This will bring up a pop-up window where you can add user names and select permissions to allow collaborators access to view, edit, or contribute to your projects.



Share project

Name	Access	Charges Allowed	
Benjamin Hitz (hitz)	Viewer		Remove
Eurie Hong (euriehong)	Admin	$	

Examples:
jsmith
user-jsmith