

ENCODE CONSORTIUM – USERS MEETING 2015

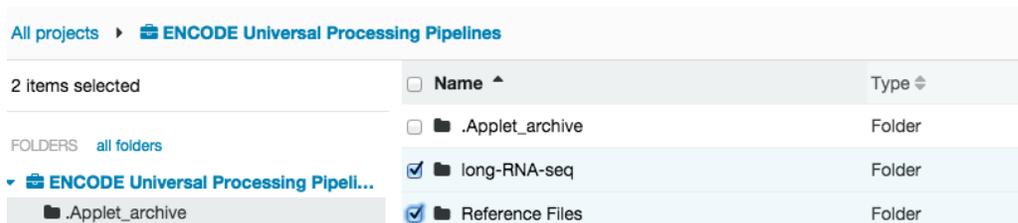
How To: Run a RNA-seq analysis pipeline on DNAnexus

Overview: In this demo, we will run the ENCODE Uniform Processing RNA-seq Pipeline on the results of an experiment using human fetal stomach tissue (<https://www.encodeproject.org/experiments/ENCSR000AFI/>) with the DNAnexus interface. This pipeline is also available for download via github at: <https://github.com/ENCODE-DCC/long-rna-seq-pipeline>

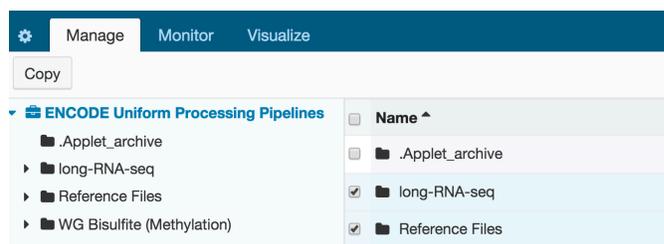
- 1) Log in to your DNAnexus account at www.dnanexus.com
- 2) Choose “ENCODE Uniform Processing Pipeline” from the “Featured Projects” list on the left hand side of the page.



- 3) Select the boxes by “long-RNA-seq” and “Reference Files” to select these files.

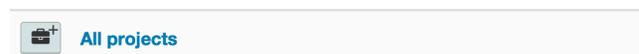


- 4) Select “Copy” to copy a version of these files to your account.

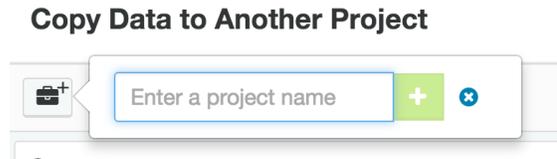


- 5) This will bring up a pop-up window that will allow you to save these files. In this demo, select the “New Project” button, which is at the top left corner. If you already have projects made, you can also select one of those here.

Copy Data to Another Project



- 6) Enter a name for your new project when prompted. Select the green “plus” sign to add when name is finished.



- 7) Select “Copy to this folder” to add the ENCODE Uniform Processing Pipeline” files to this new project.



- 8) When finished, the following pop-up window will appear:



- 9) Select the arrow at the top left of the browser window to return to the home page.



- 10) Select your project from the project management list.

ENCODE_Demo	Admin	1	348.41 GB	Jun 26, 20...
-------------	-------	---	-----------	---------------

- 11) Select the “long-RNA-seq” text.

<input type="checkbox"/> Name ^	Type ⇅
<input type="checkbox"/> <u>long-RNA-seq</u>	Folder

- 12) Select “ENCODE RNA-seq (Long) Pipeline – 1 replicate (paired-end)”

<input type="checkbox"/> Name ^	Type ⇅
<input type="checkbox"/> archive	Folder
<input type="checkbox"/> examples	Folder
<input type="checkbox"/> ENCODE RNA-Seq (Long) Pipeline - 1 replicate (paired-end)	Workflow
<input type="checkbox"/> ENCODE RNA-Seq (Long) Pipeline - 1 replicate (single-end)	Workflow

- 13) This will bring up the workflow for the ENCODE RNA-seq (Long) Pipeline using 1 paired end replicate. You will now populate the left side of the page (the orange boxes are input files) with the appropriate files. All the blue and green (outputs) boxes are auto-filled by the pipeline. The black boxes represent the various processing steps.

Run "ENCODE RNA-Seq (Long) Pipeline - 1 replicate (paired-end)" as Analysis ✕

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

- 14) If you hover over the input boxes, it will prompt you as to which file is required.

- 15) Selecting the input box will open a new window that will list possible options. There is a query window in the upper right corner of the pop-up window that can be used to filter the files by reference genome, sex, or name.

Select data for Pair 1 Reads to align (fastq.gz) input Long-RNA-Seq-align-star-pe

All projects ▸ **ENCODE_Demo**

PATTERNS [clear](#)

Files (*.fastq.gz, *.fq.gz)

FOLDERS

- ▾ ENCODE_Demo
 - ▾ long-RNA-seq
 - ▾ Reference Files

Name	Type
ENCFF646CCF_1-chr15.fastq.gz	/long-RNA-seq/exam File
ENCFF646CCF_2-chr15.fastq.gz	/long-RNA-seq/exam File
ENCFF646CCF_1-chr21.fastq.gz	/long-RNA-seq/exam File
ENCFF646CCF_1-chr21hemi.fastq.gz	/long-RNA-seq/e File
ENCFF646CCF_2-chr21.fastq.gz	/long-RNA-seq/exam File
ENCFF646CCF_2-chr21hemi.fastq.gz	/long-RNA-seq/e File

- 16) Select "ENCFF646CCF_1-chr21hemi.fastq.gz" – this will now complete the first paired end input.

17) Repeat with Pair 2

ENCFF646CCF_2-chr21hemi.fastq.gz /long-RNA-seq/e/ File

18.55 MB

Jun 24, 2015 1:29 PM



18) Add the appropriate genome reference files for each analysis step. Be sure to use the correct male or female reference genome as appropriate for the biosample (metadata is available on the ENCODE Portal - <https://www.encodeproject.org/experiments/ENCSR000AFI>) There are specific reference genome files for STAR, tophat, and RSEM.

19) Once all inputs are filled, select the black (analysis step) boxes in the center.



20) Selecting the analysis step box will open a pop-up window. You will need to provide an id in the "Identifier for biosample library" – which will be used for all the analysis steps. Select save when finished.

Configure: Long-RNA-Seq-align-star-pe Version 2.0.0 [More info about this app](#)

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

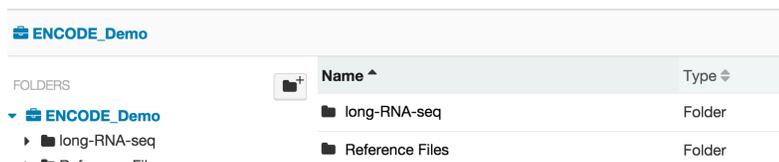
Version	2.0.0
Name	Long-RNA-Seq-align-star-pe
Output Folder	<input type="text"/>
Instance type	mem3_hdd2_x8 Select
Common	
Identifier for biosample library	Demo *
Number of threads to use	8

* Required

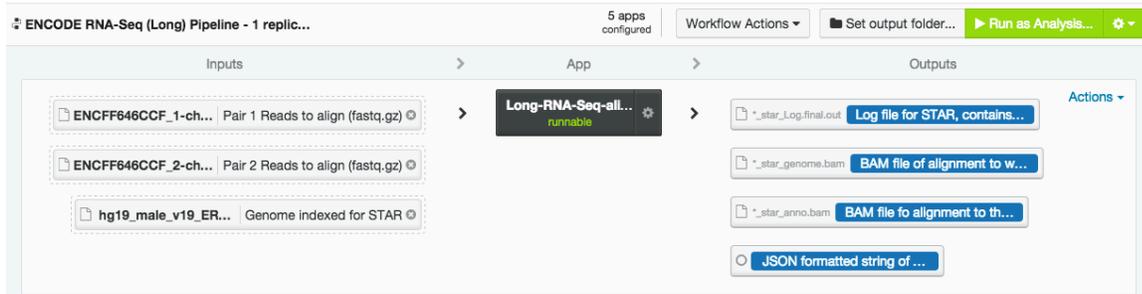
Reset to app defaults Cancel Save

21) You must specify the location of where these output files will be stored by selecting the "Set output folder button." This will open a pop-up window where the folder can be selected.

Select output folder



22) When all input files are correctly selected and all analysis steps have been configured, the text within the black box will appear green and the “Run as Analysis” box in the upper right corner will become available.



23) Starting the analysis will bring up the “Monitor” tab which will display the details of the pipeline steps as they run.

Status	Name	Executable	Launched by	Started	Duration	Price
In Progress	ENCODE RNA-Seq (Long) Pipeline - 1 replicate (paired-end)	ENCODE RNA-Seq (Long) Pip...	Eurie Hong	06/25/2015 2:59 pm	2m	—
Waiting	Long-RNA-Seq-align-star-pe	Long-RNA-Seq-align-star-pe (v2.0.0)	Eurie Hong	06/25/2015 2:59 pm	—	—
Waiting	Long-RNA-Seq-align-tophat-pe	Long-RNA-Seq-align-tophat-pe (v1.0.1)	Eurie Hong	06/25/2015 2:59 pm	—	—
Waiting on Input	Long-RNA-Seq-BAM-to-BW-stranded	Long-RNA-Seq-BAM-to-BW-stranded (Eurie Hong	06/25/2015 2:59 pm	—	—
Waiting on Input	Long-RNA-Seq-BAM-to-BW-stranded	Long-RNA-Seq-BAM-to-BW-stranded (Eurie Hong	06/25/2015 2:59 pm	—	—
Waiting on Input	Long-RNA-Seq-quantitate-RSEM	Long-RNA-Seq-quantitate-RSEM (v1.0.	Eurie Hong	06/25/2015 2:59 pm	—	—

24) If necessary, the Terminate button can be used to cancel the pipeline. Otherwise, when completed the status will change.

Status	Name	Executable	Launched by	Started	Duration	Price
Done	ENCODE RNA-Seq (Long) Pipeline - 1 replicate (paired-end)	ENCODE RNA-Seq (Long) Pip...	Eurie Hong	06/25/2015 2:59 pm	51m	\$2
Done	Long-RNA-Seq-align-star-pe	Long-RNA-Seq-align-star-pe (v2.0.0)	Eurie Hong	06/25/2015 2:59 pm	18m	\$0.43
Done	Long-RNA-Seq-align-tophat-pe	Long-RNA-Seq-align-tophat-pe (v1.0.1)	Eurie Hong	06/25/2015 2:59 pm	34m	—
Done	Long-RNA-Seq-BAM-to-BW-stranded	Long-RNA-Seq-BAM-to-BW-stranded (Eurie Hong	06/25/2015 2:59 pm	4m	—
Done	Long-RNA-Seq-BAM-to-BW-stranded	Long-RNA-Seq-BAM-to-BW-stranded (Eurie Hong	06/25/2015 2:59 pm	4m	—
Done	Long-RNA-Seq-quantitate-RSEM	Long-RNA-Seq-quantitate-RSEM (v1.0.	Eurie Hong	06/25/2015 2:59 pm	24m	—

25) The output files can be found in the specified output folder.

File Name	Size	Date
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome...	52.82 MB	Jun 24, 2015 4:26 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome...	1.81 MB	Jun 24, 2015 4:31 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome...	1.64 MB	Jun 24, 2015 4:31 PM

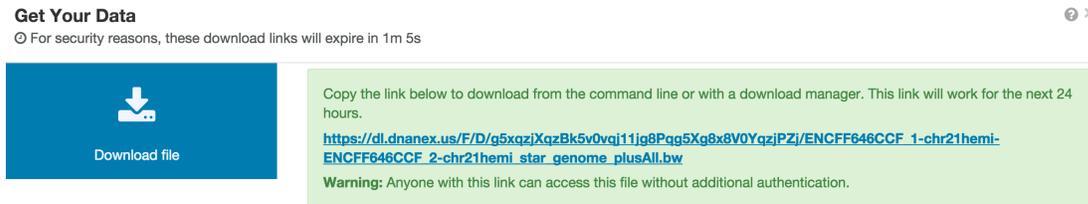
26) To visualize with a custom track hub at the UCSC Genome Browser, select the output file you are interested in utilizing. Right click and select “Download.”

File Name	Size	Date
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome_minusUniq.bw	1.6...	Jun 26, 2015 3:50 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome_plusAll.bw	1.7...	Jun 26, 2015 3:50 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome_plusUniq.bw	1.5...	Jun 26, 2015 3:50 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_Log.final.out	1.6...	Jun 26, 2015 3:45 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_tophat.bam	44...	Jun 26, 2015 4:04 PM
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_tophat_minusAll.bw	1.6...	Jun 26, 2015 4:09 PM

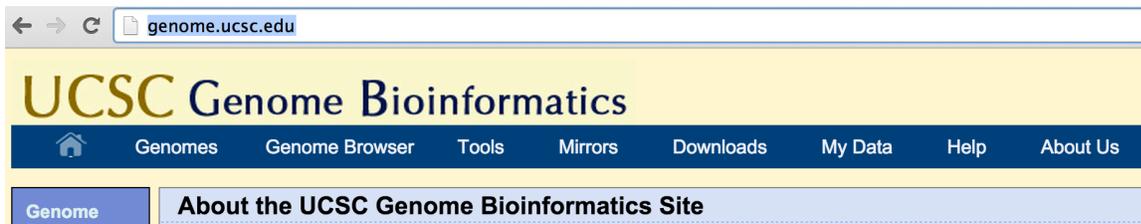
27) A new window will pop up. Select “Get URL.”



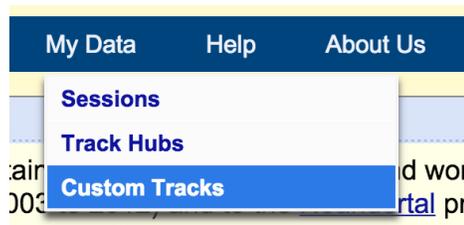
28) This will cause the window to change and a URL to the file will appear. Right click on this URL and select “Copy Link Address.”



29) Go to <http://genome.ucsc.edu/> and select “My Data” from the top options bar.



30) Select “Custom Tracks” from the options menu.



31) Paste the URL link into the text window. Be sure the reference genome is correct for this file (GrCH37 for this demo). Hit “Submit” when finished.



32) This will bring up the “Manage Custom Tracks” page. Select “Go to Genome Browser” to visualize this track.

Manage Custom Tracks

genome: Human assembly: Feb. 2009 (GRCh37/hg19) [hg19]

Name	Description	Type	Doc	delete
ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome_plusAll	ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome_plusAll	bigWig		<input type="checkbox"/>

add custom tracks
 go to genome browser
 go to table browser
 go to variant annotation integrator

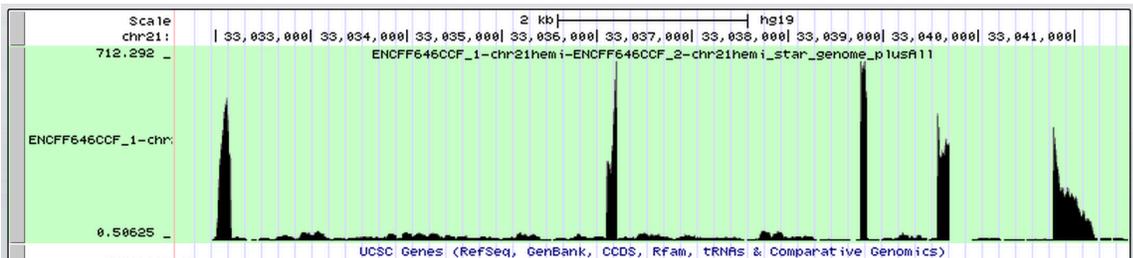
33) Scroll down and your custom track will appear on the browser page. Change visualization settings as wanted.

Custom Tracks refresh

[ENCF646CCF_1-chr21hemi-ENCF646CCF_2-chr21hemi_star_genome_plusAll](#)

dense

34) Your custom track is now visualized in the genome browser at UCSC.



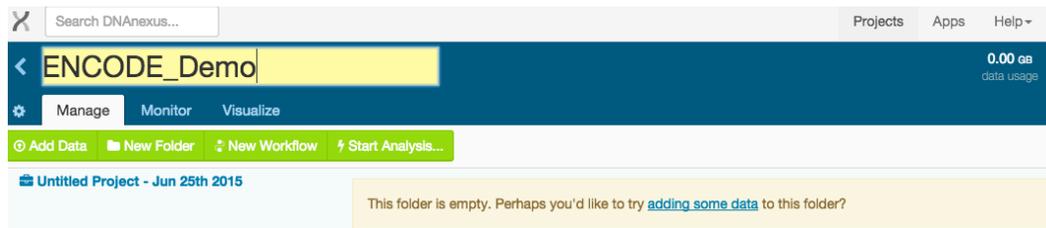
Other DNAnexus Tools:

To load data once you are in your own project

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



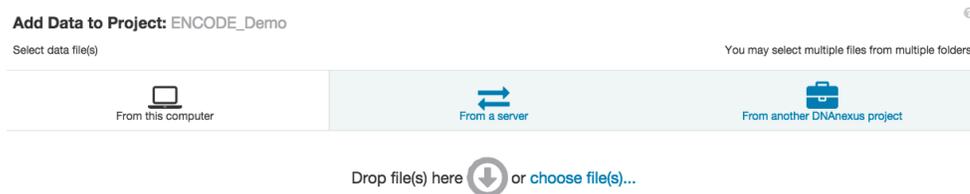
- 2) If new, name project in the upper left corner.



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



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



- 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.

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**

Close

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

ENCODE_Demo

Manage Monitor Visualize

Add Data New Folder New Workflow Start Analysis...

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
ENCF001RNE 4.78 GB	fastq	2	1	101 nt	paired-ended	2		Thomas Gingeras, CSHL	2013-07-17	pending
ENCF001f 4.8 GB				101 nt	paired-ended	1		Thomas Gingeras, CSHL	2013-07-17	pending
ENCF001f 5.15 GB				101 nt	paired-ended	2		Thomas Gingeras, CSHL	2013-07-18	pending
ENCF001f				101 nt	paired-ended	1		Thomas Gingeras, CSHL	2013-07-18	pending

Open Link in New Tab
Open Link in New Window
Open Link in Incognito Window
Save Link As...
Copy Link Address
Copy
Search Google for 'Download'
Print...

- 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.

From this computer From a server From another DNAnexus project

https://www.encodeproject.org/files/ENCF001RNE/@download/ENCF001RNE.fastq.gz

Enter a URL...

Add Data to Project: ENCODE DEMO_June24

Done

Name

https://www.encodeproject.org/files/ENCF001RNE/@download/ENCF001RNE.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.

Admin Private Share

your access access policy 2 Members

- 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	\$

Add member...

Examples:
jsmith
user-jsmith

Close