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.
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.

11) Select the “long-RNA-seq” text.

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

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.

16) Select “ENCFF646CCF_1-chr21hemi.fastq.gz” – this will now complete the first paired end input.
17) Repeat with Pair 2

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.

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.
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.

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

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

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.”
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.

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

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.

6) Choose “Add Data” to select these files.
7) When these files are uploaded, the following window will pop up.

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

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.”

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.
**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.