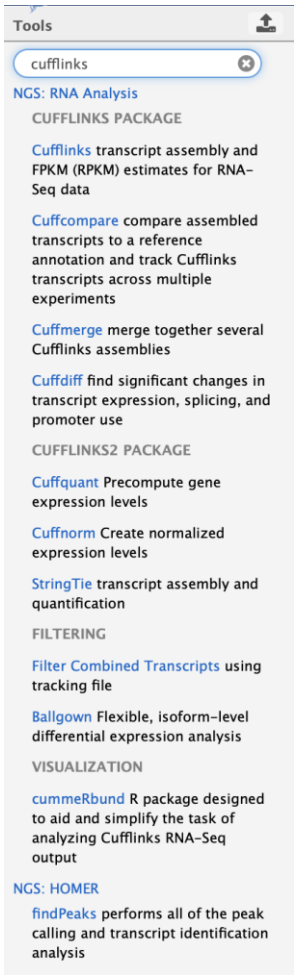
RNAseq Mapping 1a

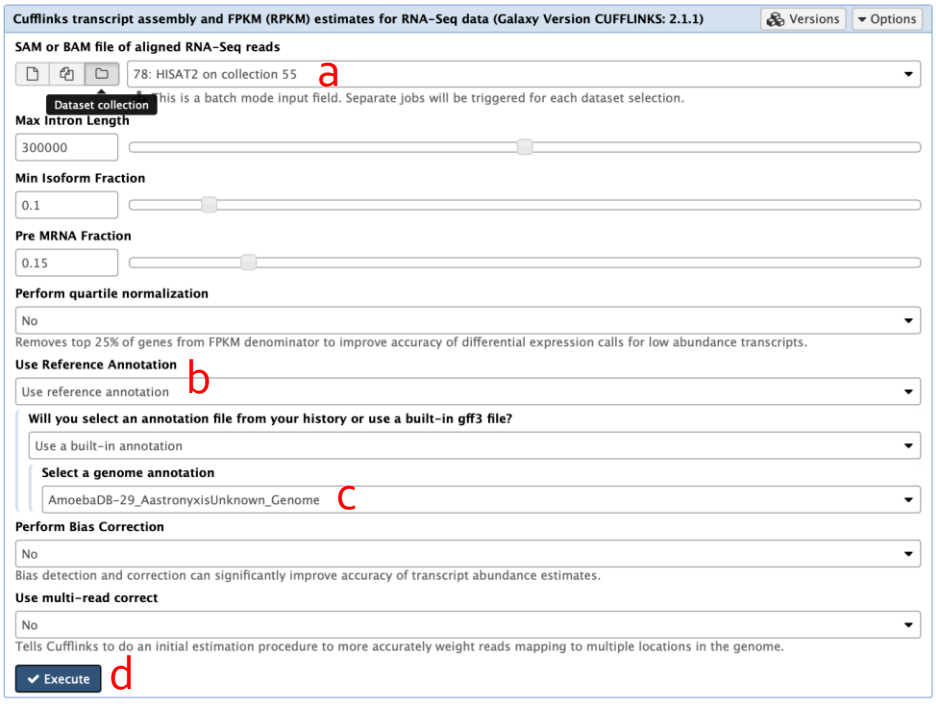
This addendum will allow you to take advantage of the EuPathDB RNAseq export tool which provides a mechanism to query your RNAseq results (FPKM values) using EuPathDB search tools.

To generate FPKM values you need a tool called Cufflinks and read alignment files – BAM files. Our workflow from yesterday generated BAM alignment files from a tool called HISAT2.

Follow these steps to generate FPKM values:

1. Find the tool called Cufflinks by typing the word cufflinks in the tool search box on the left-hand side.
2. Click on the tool to access its parameters.
3. Modify the cufflinks parameters
    - Change the input file to collection and select one of the HISAT2 collections
    - Change the Use Reference Annotation from "No" to "use reference annotation'
    - Select the appropriate reference genome from the drop down list
    - Click on execute.





After Cufflinks is done running, the next step is to take the FPKM output files from the collection outputs and put them into a single collection. Notice that cufflinks generates three types of FPKM files (or collections in this case): 1. Gene expression 2. Transcript expression 3. Assembled transcripts. We will only worry about the gene expression files for this section.

- Since we have collections of output files we will need to show all hidden files so we can generate the single collection. To do this, click on the word hidden in the upper right-hand side of the screen
- This will expose all hidden files.
- Click on the check box to perform an operation on multiple datasets (arrow in above image)
- Find all files containing the words "gene expression" and select all the cufflinks files (**NOT** the collections)
- Build a dataset list by clicking on the "for all selected" button and select "Build dataset list".
- Rename each of the datasets in the list and give this collection a meaningful name.