Print

Tip Sheet

Importing files from a shared history

- *Is there an alternative to downloading and then uploading the assigned fastq files into my Galaxy space?* It might be fastest to *import the entire history* after you login and then delete the unneeded files (by clicking the X icon of each).
- *Can't I move just a few needed files instead of an entire history?* Apparently not. However, after you import a new history into your space, you can then copy one file (or any subset) from the new history into a pre-existing history: (1) In the right History panel, click the Options gear icon and select Copy Datasets; (2) then in the middle panel, select the Source History with the desired data file, check the appropriate boxes of each desired item, choose the appropriate Destination History, and click the Copy History Items button below. After you confirm that the desired data file is in the destination history, then to save disk space you can delete the unneeded, source history.

Converting sequence files to Sanger format

- *Why can't FASTQ Quality Trimmer open the imported sequence files?* I'm not sure. Looking at the 4th (quality score) line, I saw some digits indicating the sanger format, but I first had to **convert each sequence file using FASTQ Groomer (with default settings)** before I could run FASTQ Trimmer.
    - Alternatively one could change the datatype by editing the attributes of the file to .fastqsanger
- *Which trimmer should I use?* I recommend the FASTQ Quality Trimmer, which trims each read individually, instead of trimming all reads at the same location.
- *What settings for FASTQ Quality Trimmer?* The default settings, which do a "simple" trim are a good place to start.

Mapping Processed Reads

- *Where's Tophat?* **Tophat** is in the NGS Toolbox, under NGS: RNA-seq.
- *Which reference genome?* When the Galaxy server updated its Tophat tool, it didn't include the yeast genome among its list of reference genomes. I posted a request to add this today (4/23/15), but in the meantime you will need to "**Use a genome from history**." But first, you must **first add the yeast genome sequence** to your current history. You can now find the yeast genome sequence file **in the PsAvh172 shared history** (where you previously accessed your RNA-seq files). See tip above for how to import just one file into the history that you've

been working with.

- *Is library mate-paired?* Yes. Each group is assigned a pair of files, ending with R1 and R2. These two files are mate-paired.
- *Mean Inner Distance between Mate Pairs?* Examining the RNA-seq Library Information spreadsheet (on woodle), you'll find that the average fragment size was around 380 bp in each library (ranging from 359-394 bp). FastQC indicates that the length of each read is 100 bp. Therefore, the **mean inner distance between the mate pairs is on average 180 bp**: 380 bp total - 100 bp from the right end - 100 bp from the left end.

Assembling Transcripts (with Cufflink)

- *Max Intron Length?* According to SGD/YeastMine, the longest intron is <**2500 bases**.
- *Use Reference Annotation?* Yes, "Use reference annotation as guide." The file **Yeast_allORFs.gff3** (available with the RNA seq reads) contain all the predicted mRNA sequences in yeast.
- *Improve accuracy?* There are several settings to improve accuracy, which **I recommend using**.

Merging Transcripts (with Cuffmerge)

- *What settings were used when running cuffmerge with our data?* After you import the shared history at https://usegalaxy.org/u/wmorgan /h/psavh172-merged-data--analysis, you can

  1. click on the cuffmerge dataset (#8) to reveal the format, etc., and
  2. then click the "Run this job again" icon to display the settings used for this process.

Differential Expression Analysis (with Cuffdiff)

- *What settings were used when running cuffdiff with our data?* After you import the shared history at https://usegalaxy.org/u/wmorgan /h/psavh172-merged-data--analysis, you can

  1. click on any cuffdiff dataset (#15-29) to reveal the format, etc., and
  2. then click the "Run this job again" icon to display the settings used for this process.