Transcriptome profiling exercise

Transcriptome profiling exercise

Use Binf's local galaxy server:

galaxy.bio.ku.dk

Use your previously created usernames and passwords. If none, then make them first!

Transcriptome profiling exercise

- Assignment: Use galaxy to
    1. Import data files (bam files)
    2. Assemble and quantify transcript (via Cufflinks)
    3. Make a combined transcriptome (via Cuffmerge)
    4. Make differential expression analysis (via Cuffdiff)
- And answer the questions along the way

Transcriptome profiling exercise Part 1

- You will use four .bam files for this exercise: WT1, WT2, KD1 and KD2, WT representing wild type (control) and KD is knock-down (experiment)
- BAM files are the output of Bowtie (and Tophat) meaning files that show where reads map to the genome
- These bam files are from a small loci (500 kb) on chromosome 1. They are from an experiment were HOXA1 was knocked down.

Transcriptome profiling exercise Part 1 (1 of 2)

1. Go to local BRIC galaxy: http://galaxy.bio.ku.dk/ and log in
2. Make a new history and call it something like “RNA-seq exercise”
3. Go to “Files” on Absalon -> RNA_seq folder and download the 4 .bam files to use for this exercise (KD1.bam, KD2.bam, WT1.bam and WT2.bam)

Transcriptome profiling exercise Part 1 (2 of 2)

4. Upload to Galaxy each of the 4 bam files. When uploading the first specify it is for hg19 (it will do so for the others automatically) Leave everything else at default.

- Hint: Galaxy allows you to queue tasks, meaning you can queue the upload of all 4 files

- Question: Why can't you use the "view data" (eye icon) for the bam files?
  Hint: What are bam files? they are binary format

Transcriptome profiling exercise Part 2 (1 of 3)

- That should result in something like this in the data menu (right side):

  - The order and numbers does not matter
- Question: Why can't you use the "view data" (eye icon) for the bam files?
  Hint: What are bam files? information of reads alignment to the genome

Transcriptome profiling exercise Part 2 (2 of 3)

- Find the “NGS: RNA Analysis” section in the tool menu - here we have everything we need for RNA-seq analysis
- Remember! Read the whole documentation page before using any tool!

Transcriptome profiling exercise Part 2 (3 of 3)

- For each of the 4 bam files assemble the transcripts using the “cufflinks” tool with default parameters (run cufflinks 4 times)
- Hint: We only need the data called “assembled transcripts” meaning you can delete the other files (and preserve the overview)
- Question: For the cufflinks run on WT1.bam, how many transcripts found consists of 1 exon (manually inspect the files)? [5]
  Hint: The “view data” button might be usefull
  Hint: Look up a GTF file online

Transcriptome profiling exercise Part 3 (1 of 1)

- Using the input of all 4 cufflinks runs (the “assembled transcripts”) create a combined transcriptome with the “cuffmerge” tool. Use default parameters
- Question: What is a GTF file (google it)

The GFF (General Feature Format) format consists of one line per feature, each containing 9 columns of data, plus optional track definition lines

- Question: How would you obtain the GTF file, containing the RefSeq genes or UCSC knownGenes (both are reference transcriptome/annotation), required if you want to use the “Use Reference Annotation” parameter?

Transcriptome profiling exercise Part 4 (1 of 2)

- Use the Cuffdiff tool, with default parameters, to make de differential expression analysis. Use all 4 bam files and the combined transcriptome create via cuffmerge
- Today we only need “transcript differential expression testing” and “gene differential expression testing”.

Transcriptome profiling exercise Part 4 (2 of 2)

- How many transcripts were found in the combined transcriptome? How many genes?

10genes 5 transcript

- How many significantly differentially expressed genes were found? Transcripts?
- What does the ratio of significant gene/ transcripts tell you?