Edit README.md
gohr authored just now

README.md 4.58 KiB

Steps of Galaxy course

- Discuss structure and content of decompressed live-coding data
  - Including
- Upload GTF
- Rename to `yeast_genes_R64.gtf`
- Upload FASTA.gz
- Rename to `yeast_genomeseqs_R64.fasta.gz`
- Upload PE reads FASTQ.gzs
  - Announce clearly that file type must be fastqsanger.gz
- Show one FASTQ file and GTF and FASTA
- Create collection of paired FASTQ.gzs
- Rename to `2017_wilkins_orig_PE_reads`
- Show and discuss `2017_wilkins_orig_PE_reads`
- Run cutadapt on `2017_wilkins_orig_PE_reads`
  - PE mode
  - Illumina universal adapter seq AGATCGGAAGAG
  - min read length 17
  - Output: Two collections R1 and R2
  - Refer to cutadapt manual
- Show and discuss output; Show trimmed reads
- Rename output to `2017_wilkins_trimmed_reads_1` and `2017_wilkins_trimmed_reads_2`
- Run FASTQC on `2017_wilkins_trimmed_reads_1`
  - Output for each run: Webpages and text output
  - Refer to FASTQC manual
- Rename to `FastQC_trimmed_reads_1`, `FastQC_trimmed_reads_1_txt`
- Show and discuss `FastQC_trimmed_reads_1_txt`
- Run FASTQC on `2017_wilkins_trimmed_reads_2`
  - Output for each run: Webpages and text output
  - Refer to FASTQC manual
- Rename to `FastQC_trimmed_reads_2`, `FastQC_trimmed_reads_2_txt`
- Show and discuss `FastQC_trimmed_reads_2_txt`
- Run STAR
  - Paired-end as individual data sets
    - Forward reads: Collection `2017_wilkins_trimmed_reads_1`
    - Reverse reads: Collection `2017_wilkins_trimmed_reads_2`
  - Use reference genome from history and create temporary index
    - Choose `yeast_genomeseqs_R64.fasta.gz`
    - Choose `yeast_genes_R64.gtf`
  - Count number of reads per gene: Yes
  - Rest keep default params
  - Refer to STAR manual
- Rename output to `2017_wilkins_bams`, `2017_wilkins_reads_per_gene`
- Check `STAR_logs` and point out high fraction of uniquely mapping reads and low fraction of un-mapped reads
- Check `STAR_bams`
  - Can vbe downloaded and visualized in IGV
- Check `2017_wilkins_reads_per_gene`
  - Discuss structure of table and what info we need for DESeq2
  - Reformat tables with tools `Select last (tail) everything from line 5` and `cut c1,c2`
  - Rename output to `2017_wilkins_reads_per_gene_unstranded`
  - Label files from collection `2017_wilkins_reads_per_gene_unstranded` by their file names so we find them later easily
  - Delete intermediate output from tool `tail`
- Run DESeq2
  - Choose `Select dataset per level`
  - Define factor Treatment with two levels `amb` and `ctr`
    - Choose amb as first level as log2FCs will be level1 vs level2

    - Search hidden files for files in collection `2017_wilkins_reads_per_gene_unstranded`
    - Assign read-count files to levels
    - Output: Choose plots and rLog normalized
- Go through results and explain
- Run Annotate DESeq2 result tables
    - Discuss result
    - Add column names
        1. Add new files via pasting content and selecting type tabular and name it Header
        2. Contents commas need to be changed to tabs in text editor first: GeneID,Base mean,log2(FC),StdErr,Wald-Stats,P-value,P-adj,Chromosome,Start,End,Strand,Feature,Gene name
        3. Run concatenate datasets tail-to-head and choose Header and DESeq2 annotated table
        4. Rename resulting table as `Annotated DESeq2 result`
- Explain goal of next steps: We want do create a heatmap with all samples and diff. expressed genes. Therefore, we need to define diff. expressed genes and subset the table with normalized read counts to these DEGs.
- Filter data on any column
    - `abs(c7)<0.01`: c7 refers to column 7 and we select rows where that entry (padj) is smaller than 0.01
    - Number of header lines to skip: `1`
    - Rename output `DEGs padj 0.01`
- Run Join Two Datasets side by side:
    - File 1: `Normalized read counts from DESeq2`
    - Using column `1`
    - File 2: `DEGs padj 0.01`
    - Using column `1`
    - Choose `Keep header lines: Yes`
- Run Cut Columns From Table
    - File: Output from previous step
    - c1-c7 (first column with gene_id and then per sample one column with normalized read counts)
    - tab
    - Rename output `NormCounts DEGs padj 0.01`
- Run Heatmap2
    - File `NormCounts DEGs padj 0.01`
    - Data as is
    - Z-normalization: yes on rows only
    - Cluster: rows and columns only
    - Labeling: columns only
    - Color map: 3 colors
    - File type: PNG
    - Choose PDF for high-resolution, must be downloaded