

RNA_sequencing analysis

Meghan Pepler

27/04/2021

RNA-sequencing analysis of two *S. venezuelae* strains:

1) RNase III mutant: Δrnc

2) RNase J mutant: Δrnj

These two ribonuclease mutants have interesting phenotypes in terms of their development and secondary metabolism. We are primarily interested finding differentially expressed regions in either ribonuclease mutant that are downstream of an orphan TSS.

Later we will also investigate significantly differentially expressed genes in the RNase III mutant that also have antisense TSSs (and thus are potentially regulated by both small RNAs and the dsRNA-specific catalytic activity of RNase III)

First step is **quality checking** reads!

FastQC

```
#using the fastqr program to quality check the sequences
#For the sake of being able to properly knit the markdown some functions are commented out

#install.packages("fastqcr")
library(fastqcr)
fastqc_install(dest.dir = "~/Desktop/RNA_seq/")

#run the fastQC tool on all .fastq files
 #(it will automatically run on every .fastq file in the directory you call)

fastqc(fq.dir = "~/Desktop/RNA_seq/raw/RNaseIII_raw",
      qc.dir = "~/Desktop/RNA_seq/FastQC_output/rnc",
      fastqc.path = "~/Desktop/RNA_seq/FastQC/fastqc")

fastqc(fq.dir = "~/Desktop/RNA_seq/raw/RNaseJ_raw",
      qc.dir = "~/Desktop/RNA_seq/FastQC_output/rnj",
      fastqc.path = "~/Desktop/RNA_seq/FastQC/fastqc")

fastqc(fq.dir = "~/Desktop/RNA_seq/raw/WT_raw/",
      qc.dir = "~/Desktop/RNA_seq/FastQC_output/WT",
      fastqc.path = "~/Desktop/RNA_seq/FastQC/fastqc")
```

If you run this, you can go into the output folder on your local computer and read through the html FastQC outputs. I will also include hyperlinks for the next FastQC output!

From the output you can see the reads still contain adapters. Gotta trim them!

Trim using RBowtie2 remove_adapter() function

Trimming reads from RNase III mutant (gene: rnc) - 10hr

```
library(Rbowtie2)

#remove adapter rnc_A_10h_R1
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_A_10hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGAGAGTACATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/rnc/rnc_A_10h_R1")

#remove adapter rnc_A_10h_R2
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_A_10hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGCTTCCAGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/rnc/rnc_A_10h_R2")

#remove adapter rnc_B_10h_R1
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_B_10hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACTCTACGCAATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/rnc/rnc_B_10h_R1")

#remove adapter rnc_B_10h_R2
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_B_10hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGGCTATTGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/rnc/rnc_B_10h_R2")
```

Trimming reads from RNase III mutant (gene: rnc) - 14hr

```
#remove adapter rnc_A_14h_R1
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_A_14hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/rnc/rnc_A_14h_R1")

#remove adapter rnc_A_14h_R2
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_A_14hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGAGATACGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/rnc/rnc_A_14h_R2")

#remove adapter rnc_B_14h_R1
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_B_14hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGACCTTGATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACAGACCTTGATCTCGTAT",
  output1 = "trimmed/rnc/rnc_B_14h_R1")

#remove adapter rnc_B_14h_R2
```

```
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_B_14hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGCACGTAAGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACAGACCTTGATCTCGTAT",
  output1 = "trimmed/rnc/rnc_B_14h_R2")
```

Trimming reads from RNase III mutant (gene: rnc) - 20hr

```
#remove adapter rnc_A_20h_R1
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_A_20hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGAATCCGTATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCATGAGCAATCTCGTAT",
  output1 = "trimmed/rnc/rnc_A_20h_R1")

#remove adapter rnc_A_20h_R2
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_A_20hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGGAGTTGGTGTAGATCT",
  adapter2 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGCAAGATCGTGTAGAT",
  output1 = "trimmed/rnc/rnc_A_20h_R2")

#remove adapter rnc_B_20h_R1
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_B_20hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACCATGAGCAATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCATGAGCAATCTCGTAT",
  output1 = "trimmed/rnc/rnc_B_20h_R1")

#remove adapter rnc_B_20h_R2
remove_adapters(file1 = "raw/RNaseIII_raw/rnc_B_20hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGCAAGATCGTGTAGATCT",
  adapter2 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGCAAGATCGTGTAGAT",
  output1 = "trimmed/rnc/rnc_B_20h_R2")
```

Trimming reads from RNase J mutant (gene: rnj) - 12hr

```
library(Rbowtie2)

#remove adapter rnj_A_12h_R1
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_A_12hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_A_12h_R1")

#remove adapter rnj_A_12h_R2
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_A_12hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGTTTCGAGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_A_12h_R2")

#remove adapter rnj_B_12h_R1
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_B_12hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTCAGAAGATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCTCAGAAGATCTCGTAT",
  output1 = "trimmed/rnj/rnj_B_12h_R1")
```

```
#remove adapter rnj_B_12h_R2
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_B_12hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTAGGGAAAGAGTGTAACTTGGCGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCTCAGAAGATCTCGTAT",
  output1 = "trimmed/rnj/rnj_B_12h_R2")
```

Trimming reads from RNase J mutant (gene: rnj) - 18hr

```
#remove adapter rnj_A_18h_R1
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_A_18hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCGTTACATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_A_18h_R1")

#remove adapter rnj_A_18h_R2
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_A_18hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTAGGGAAAGAGTGTGCTTAGCTGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_A_18h_R2")

#remove adapter rnj_B_18h_R1
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_B_18hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTAACCGAATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_B_18h_R1")

#remove adapter rnj_B_18h_R2
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_B_18hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTAGGGAAAGAGTGTGGTGTCTTGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_B_18h_R2")
```

Trimming reads from RNase J mutant (gene: rnj) - 24hr

```
#remove adapter rnj_A_24h_R1
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_A_24hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTAGGACATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_A_24h_R1")

#remove adapter rnj_A_24h_R2
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_A_24hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTAGGGAAAGAGTGTCAAGTGAAGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_A_24h_R2")

#remove adapter rnj_B_24h_R1
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_B_24hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCTGACCATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_B_24h_R1")

#remove adapter rnj_B_24h_R2
```

```
remove_adapters(file1 = "raw/RNaseJ_raw/rnj_B_24hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAAAGTCGAGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATTCATCTCGTAT",
  output1 = "trimmed/rnj/rnj_B_24h_R2")
```

Trimming reads from wild type (WT) - 10hr

```
#remove adapter WT_A_10h_R1
remove_adapters(file1 = "raw/WT_raw/WT_A_10hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACTCGGATTCATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/WT/WT_A_10h_R1")

#remove adapter WT_A_10h_R2
remove_adapters(file1 = "raw/WT_raw/WT_A_10hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGTTGCGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/WT/WT_A_10h_R2")

#remove adapter WT_B_10h_R1
remove_adapters(file1 = "raw/WT_raw/WT_B_10hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTGTACCAATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/WT/WT_B_10h_R1")

#remove adapter WT_B_10h_R2
remove_adapters(file1 = "raw/WT_raw/WT_B_10hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGTCTGTGGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/WT/WT_B_10h_R2")
```

Trimming reads from wild type (WT) - 14hr

```
#remove adapter WT_A_14h_R1
remove_adapters(file1 = "raw/WT_raw/WT_A_14hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCCTAAGATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/WT/WT_A_14h_R1")

#remove adapter WT_A_14h_R2
remove_adapters(file1 = "raw/WT_raw/WT_A_14hr_R2.fastq",
  adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGTTAGTCTGTGTAGATCT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACCAAGGTACATCTCGTAT",
  output1 = "trimmed/WT/WT_A_14h_R2")

#remove adapter WT_B_14h_R1
remove_adapters(file1 = "raw/WT_raw/WT_B_14hr_R1.fastq",
  adapter1 = "GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCGTTAGAATCTCGTAT",
  adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCGTTAGAATCTCGTAT",
  output1 = "trimmed/WT/WT_B_14h_R1")

#remove adapter WT_B_14h_R2
remove_adapters(file1 = "raw/WT_raw/WT_B_14hr_R2.fastq",
```

```

adapter1 = "GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTGTATGCTGGTGTAGATCT",
adapter2 = "GCTCGGAAGAGCACACGTCTGAACTCCAGTCACGCGTTAGAATCTCGTAT",
output1 = "trimmed/WT/WT_B_14h_R2")

```

Trimming reads from wild type (WT) - 20hr

Note that for this sample it is not paired-end. The WT data came from another RNA-sequencing experiment; we were not able to isolated RNA from WT cells during sporulation during the RNase mutant RNA-isolation experiments. The wild type strain is the same as what was sequenced previously.

```

#remove adapter WT_A_20h
remove_adapters(file1 = "raw/WT_raw/WT_A_20hr.fastq",
                 adapter1 = "TTTTCAAGCAGAAGACGGCATACGAGATATCCACTCGTGAAGTTCAGACGTGTGCTCTTCCGAT",
                 adapter2 = "TTTTTCAAGCAGAAGACGGCATACGAGATATCCACTCGTGAAGTTCAGACGTGTGCTCTTCCGAT",
                 output1 = "trimmed/WT/WT_A_20h")

#remove adapter WT_B_20h
remove_adapters(file1 = "raw/WT_raw/WT_B_20hr.fastq",
                 adapter1 = "TTTTTCAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT",
                 adapter2 = "GACTGGAGTTCAGACGTGTGCTCTTCCGATC",
                 output1 = "trimmed/WT/WT_B_20h")

```

FastQC again!

Great! Time to check the quality of the fastq files again!

```

#run the fastQC tool on all .fastq files
 #(it will automatically run on every .fastq file in the directory you call)

fastqc(fq.dir = "~/Desktop/RNA_seq/trimmed/rnc/",
       qc.dir = "~/Desktop/RNA_seq/FastQC_output_2/rnc",
       fastqc.path = "~/Desktop/RNA_seq/FastQC/fastqc")

fastqc(fq.dir = "~/Desktop/RNA_seq/trimmed/rnj/",
       qc.dir = "~/Desktop/RNA_seq/FastQC_output_2/rnj",
       fastqc.path = "~/Desktop/RNA_seq/FastQC/fastqc")

fastqc(fq.dir = "~/Desktop/RNA_seq/trimmed/WT/",
       qc.dir = "~/Desktop/RNA_seq/FastQC_output_2/WT",
       fastqc.path = "~/Desktop/RNA_seq/FastQC/fastqc")

```

Links to the output FastQC files are here:

RNaseIII

RNaseIII 10h files rnc 10h R1 - A rnc 10h R2 - A rnc 10h R1 - B rnc 10h R2 - B

RNaseIII 14h files rnc 14h R1 - A rnc 14h R2 - A rnc 14h R1 - B rnc 14h R2 - B

RNaseIII 20h files rnc 20h R1 - A rnc 20h R2 - A rnc 20h R1 - B rnc 20h R2 - B

RNaseJ

RNaseJ 12h files rnj 12h R1 - A rnj 12h R2 - A rnj 12h R1 - B rnj 12h R2 - B

RNaseJ 18h files rnj 18h R1 - A rnj 18h R2 - A rnj 18h R1 - B rnj 18h R2 - B

RNaseJ 24h files rnj 24h R1 - A rnj 24h R2 - A rnj 24h R1 - B rnj 24h R2 - B

WT

WT 10h files rnc 10h R1 - A rnc 10h R2 - A rnc 10h R1 - B rnc 10h R2 - B

WT 14h files rnc 14h R1 - A rnc 14h R2 - A rnc 14h R1 - B rnc 14h R2 - B

WT 20h files rnc 20h R1 - A rnc 20h R2 - A rnc 20h R1 - B rnc 20h R2 - B

We can see that the adapters have been successfully removed, but there still seems to be some high duplication levels. These may be genuine duplications - particularly as they seem to be more prominent in later time-points (late fragmentation into sporulation). During sporulation the transcriptional activity of cells tends to decrease significantly - so the RNA that is isolated and sequenced may align to a smaller proportion of genes and thus have a greater chance of sequence duplication.

We will continue with the analysis!

The next step will be to **align these reads to the genome using Bowtie2!**

Align reads to genome using Bowtie2

Bowtie2 rnc

```
library(Rsamtools)

bowtie2_build(references="~/Desktop/RNA_seq/ref/vnz_genome.fasta", bt2Index = "~/Desktop/RNA_seq/ref/venezuelae_bowtie")

#rnc 10h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnc_10h_A",
        seq1 = "raw/RNaseIII_raw/rnc_A_10hr_R1.fastq",
        seq2 = "raw/RNaseIII_raw/rnc_A_10hr_R2.fastq")

#rnc 10h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnc_10h_B",
        seq1 = "raw/RNaseIII_raw/rnc_B_10hr_R1.fastq",
        seq2 = "raw/RNaseIII_raw/rnc_B_10hr_R2.fastq")

#rnc 14h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnc_14h_A",
        seq1 = "raw/RNaseIII_raw/rnc_A_14hr_R1.fastq",
        seq2 = "raw/RNaseIII_raw/rnc_A_14hr_R2.fastq")

#rnc 14h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnc_14h_B",
        seq1 = "raw/RNaseIII_raw/rnc_B_14hr_R1.fastq",
        seq2 = "raw/RNaseIII_raw/rnc_B_14hr_R2.fastq")

#now since the WT data for this timepoint is single-end (not paired-end) we will only call one of the s

#rnc 20h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
```

```

        samOutput = "sam_out/rnc_20h_A",
        seq1 = "raw/RNaseIII_raw/rnc_A_20hr_R1.fastq",
        seq2 = NULL)
#rnc 20h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnc_20h_B",
        seq1 = "raw/RNaseIII_raw/rnc_B_20hr_R1.fastq",
        seq2 = NULL)

```

Bowtie2 rnj

```

#rnj 12h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnj_12h_A",
        seq1 = "raw/RNaseJ_raw/rnj_A_12hr_R1.fastq",
        seq2 = "raw/RNaseJ_raw/rnj_A_12hr_R2.fastq")
#rnj 12h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnj_12h_B",
        seq1 = "raw/RNaseJ_raw/rnj_B_12hr_R1.fastq",
        seq2 = "raw/RNaseJ_raw/rnj_B_12hr_R2.fastq")

#rnj 18h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnj_18h_A",
        seq1 = "raw/RNaseJ_raw/rnj_A_18hr_R1.fastq",
        seq2 = "raw/RNaseJ_raw/rnj_A_18hr_R2.fastq")
#rnj 18h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnj_18h_B",
        seq1 = "raw/RNaseJ_raw/rnj_B_18hr_R1.fastq",
        seq2 = "raw/RNaseJ_raw/rnj_B_18hr_R2.fastq")

```

#now since the WT data for this timepoint is single-end (not paired-end) we will only call one of the s

```

#rnj 24h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnj_24h_A",
        seq1 = "raw/RNaseJ_raw/rnj_A_24hr_R1.fastq",
        seq2 = NULL)
#rnj 24h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/rnj_24h_B",
        seq1 = "raw/RNaseJ_raw/rnj_B_24hr_R1.fastq",
        seq2 = NULL)

```

Bowtie2 WT

```

#WT 12h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/WT_10h_A",
        seq1 = "raw/WT_raw/WT_A_10hr_R1.fastq",
        seq2 = "raw/WT_raw/WT_A_10hr_R2.fastq")

```



```

#WT 12h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/WT_10h_B",
        seq1 = "raw/WT_raw/WT_B_10hr_R1.fastq",
        seq2 = "raw/WT_raw/WT_B_10hr_R2.fastq")

#WT 18h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/WT_14h_A",
        seq1 = "raw/WT_raw/WT_A_14hr_R1.fastq",
        seq2 = "raw/WT_raw/WT_A_14hr_R2.fastq")

#WT 18h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/WT_14h_B",
        seq1 = "raw/WT_raw/WT_B_14hr_R1.fastq",
        seq2 = "raw/WT_raw/WT_B_14hr_R2.fastq")

#WT 24h A
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/WT_20h_A",
        seq1 = "raw/WT_raw/WT_A_20hr.fastq",
        seq2 = NULL)

#WT 24h B
bowtie2(bt2Index = "ref/venezuelae_bowtie",
        samOutput = "sam_out/WT_20h_B",
        seq1 = "raw/WT_raw/WT_B_20hr.fastq",
        seq2 = NULL)

```

Pre-processing files for counts and differential analysis

Convert sam to bam

The bowtie2 function outputs .sam files which need to be converted to .bam files for further analysis:

```

#rnc 10h
asBam(file = "sam_out/rnc_10h_A", destination = "bam_out/rnc_10h_A")
asBam(file = "sam_out/rnc_10h_B", destination = "bam_out/rnc_10h_B")

#rnc 14h
asBam(file = "sam_out/rnc_14h_A", destination = "bam_out/rnc_14h_A")
asBam(file = "sam_out/rnc_14h_B", destination = "bam_out/rnc_14h_B")

#rnc 20h
asBam(file = "sam_out/rnc_20h_A", destination = "bam_out/rnc_20h_A")
asBam(file = "sam_out/rnc_20h_B", destination = "bam_out/rnc_20h_B")

#rnj 12h
asBam(file = "sam_out/rnj_12h_A", destination = "bam_out/rnj_12h_A")
asBam(file = "sam_out/rnj_12h_B", destination = "bam_out/rnj_12h_B")

#rnj 18h
asBam(file = "sam_out/rnj_18h_A", destination = "bam_out/rnj_18h_A")
asBam(file = "sam_out/rnj_18h_B", destination = "bam_out/rnj_18h_B")

#rnj 24h
asBam(file = "sam_out/rnj_24h_A", destination = "bam_out/rnj_24h_A")
asBam(file = "sam_out/rnj_24h_B", destination = "bam_out/rnj_24h_B")

```

```

#WT 10h
asBam(file = "sam_out/WT_10h_A", destination = "bam_out/WT_10h_A")
asBam(file = "sam_out/WT_10h_B", destination = "bam_out/WT_10h_B")
#WT 14h
asBam(file = "sam_out/WT_14h_A", destination = "bam_out/WT_14h_A")
asBam(file = "sam_out/WT_14h_B", destination = "bam_out/WT_14h_B")
#WT 20h
asBam(file = "sam_out/WT_20h_A", destination = "bam_out/WT_20h_A")
asBam(file = "sam_out/WT_20h_B", destination = "bam_out/WT_20h_B")

```

Sort bam files

The .bam files need to then be sorted (function sortBam)

```

#sort (by leftmost coordinates)
#rnc 10h
sortBam(file = "bam_out/rnc_10h_A.bam", destination = "sort_bam/rnc_10h_A")
sortBam(file = "bam_out/rnc_10h_B.bam", destination = "sort_bam/rnc_10h_B")
#rnc 14h
sortBam(file = "bam_out/rnc_14h_A.bam", destination = "sort_bam/rnc_14h_A")
sortBam(file = "bam_out/rnc_14h_B.bam", destination = "sort_bam/rnc_14h_B")
#rnc 20h
sortBam(file = "bam_out/rnc_20h_A.bam", destination = "sort_bam/rnc_20h_A")
sortBam(file = "bam_out/rnc_20h_B.bam", destination = "sort_bam/rnc_20h_B")

#rnj 12h
sortBam(file = "bam_out/rnj_12h_A.bam", destination = "sort_bam/rnj_12h_A")
sortBam(file = "bam_out/rnj_12h_B.bam", destination = "sort_bam/rnj_12h_B")
#rnj 18h
sortBam(file = "bam_out/rnj_18h_A.bam", destination = "sort_bam/rnj_18h_A")
sortBam(file = "bam_out/rnj_18h_B.bam", destination = "sort_bam/rnj_18h_B")
#rnj 24h
sortBam(file = "bam_out/rnj_24h_A.bam", destination = "sort_bam/rnj_24h_A")
sortBam(file = "bam_out/rnj_24h_B.bam", destination = "sort_bam/rnj_24h_B")

#WT 10h
sortBam(file = "bam_out/WT_10h_A.bam", destination = "sort_bam/WT_10h_A")
sortBam(file = "bam_out/WT_10h_B.bam", destination = "sort_bam/WT_10h_B")
#WT 14h
sortBam(file = "bam_out/WT_14h_A.bam", destination = "sort_bam/WT_14h_A")
sortBam(file = "bam_out/WT_14h_B.bam", destination = "sort_bam/WT_14h_B")
#WT 20h
sortBam(file = "bam_out/WT_20h_A.bam", destination = "sort_bam/WT_20h_A")
sortBam(file = "bam_out/WT_20h_B.bam", destination = "sort_bam/WT_20h_B")

```

Index bam files

```

library(Rsamtools)

#rnc 10h
indexBam("bam_out/rnc_10h_A.bam", "sort_bam/rnc_10h_A.bam")
indexBam("bam_out/rnc_10h_B.bam", "sort_bam/rnc_10h_B.bam")
#rnc 14h

```

```

indexBam("bam_out/rnc_14h_A.bam", "sort_bam/rnc_14h_A.bam")
indexBam("bam_out/rnc_14h_B.bam", "sort_bam/rnc_14h_B.bam")
#rnc 20h
indexBam("bam_out/rnc_20h_A.bam", "sort_bam/rnc_20h_A.bam")
indexBam("bam_out/rnc_20h_B.bam", "sort_bam/rnc_20h_B.bam")

#rnj 12h
indexBam("bam_out/rnj_12h_A.bam", "sort_bam/rnj_12h_A.bam")
indexBam("bam_out/rnj_12h_B.bam", "sort_bam/rnj_12h_B.bam")
#rnj 18h
indexBam("bam_out/rnj_18h_A.bam", "sort_bam/rnj_18h_A.bam")
indexBam("bam_out/rnj_18h_B.bam", "sort_bam/rnj_18h_B.bam")
#rnj 24h
indexBam("bam_out/rnj_24h_A.bam", "sort_bam/rnj_24h_A.bam")
indexBam("bam_out/rnj_24h_B.bam", "sort_bam/rnj_24h_B.bam")

#WT 10h
indexBam("bam_out/WT_10h_A.bam", "sort_bam/WT_10h_A.bam")
indexBam("bam_out/WT_10h_B.bam", "sort_bam/WT_10h_B.bam")
#WT 14h
indexBam("bam_out/WT_14h_A.bam", "sort_bam/WT_14h_A.bam")
indexBam("bam_out/WT_14h_B.bam", "sort_bam/WT_14h_B.bam")
#WT 20h
indexBam("bam_out/WT_20h_A.bam", "sort_bam/WT_20h_A.bam")
indexBam("bam_out/WT_20h_B.bam", "sort_bam/WT_20h_B.bam")

```

filter unmapped reads from rnc

```

#filter out unmapped sequences
#rnc 10h
filterBam(file = "sort_bam/rnc_10h_A.bam",
          index = "bam_out/rnc_10h_A.bam.bai",
          destination = "filter_bam/rnc_10h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/rnc_10h_B.bam",
          index = "bam_out/rnc_10h_B.bam.bai",
          destination = "filter_bam/rnc_10h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

#rnc 14h
filterBam(file = "sort_bam/rnc_14h_A.bam",
          index = "bam_out/rnc_14h_A.bam.bai",
          destination = "filter_bam/rnc_14h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/rnc_14h_B.bam",
          index = "bam_out/rnc_14h_B.bam.bai",
          destination = "filter_bam/rnc_14h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

```

```

#rnc 20h
filterBam(file = "sort_bam/rnc_20h_A.bam",
          index = "bam_out/rnc_20h_A.bam.bai",
          destination = "filter_bam/rnc_20h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/rnc_20h_B.bam",
          index = "bam_out/rnc_20h_B.bam.bai",
          destination = "filter_bam/rnc_20h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

```

filter unmapped reads from rnj

```

#filter out unmapped sequences
#rnj 12h
filterBam(file = "sort_bam/rnj_12h_A.bam",
          index = "bam_out/rnj_12h_A.bam.bai",
          destination = "filter_bam/rnj_12h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/rnj_12h_B.bam",
          index = "bam_out/rnj_12h_B.bam.bai",
          destination = "filter_bam/rnj_12h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

#rnj 18h
filterBam(file = "sort_bam/rnj_18h_A.bam",
          index = "bam_out/rnj_18h_A.bam.bai",
          destination = "filter_bam/rnj_18h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/rnj_18h_B.bam",
          index = "bam_out/rnj_18h_B.bam.bai",
          destination = "filter_bam/rnj_18h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

#rnj 24h
filterBam(file = "sort_bam/rnj_24h_A.bam",
          index = "bam_out/rnj_24h_A.bam.bai",
          destination = "filter_bam/rnj_24h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/rnj_24h_B.bam",
          index = "bam_out/rnj_24h_B.bam.bai",
          destination = "filter_bam/rnj_24h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

```

filter unmapped reads from WT

```

#filter out unmapped sequences
#WT 10h
filterBam(file = "sort_bam/WT_10h_A.bam",
          index = "bam_out/WT_10h_A.bam.bai",
          destination = "filter_bam/WT_10h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/WT_10h_B.bam",
          index = "bam_out/WT_10h_B.bam.bai",
          destination = "filter_bam/WT_10h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

#WT 14h
filterBam(file = "sort_bam/WT_14h_A.bam",
          index = "bam_out/WT_14h_A.bam.bai",
          destination = "filter_bam/WT_14h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/WT_14h_B.bam",
          index = "bam_out/WT_14h_B.bam.bai",
          destination = "filter_bam/WT_14h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

#WT 20h
filterBam(file = "sort_bam/WT_20h_A.bam",
          index = "bam_out/WT_20h_A.bam.bai",
          destination = "filter_bam/WT_20h_A.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))
filterBam(file = "sort_bam/WT_20h_B.bam",
          index = "bam_out/WT_20h_B.bam.bai",
          destination = "filter_bam/WT_20h_B.bam",
          param = ScanBamParam(flag=scanBamFlag(isUnmappedQuery=FALSE),
                                what="seq"))

```

Lets make some counts tables!

We will be using `FeatureCounts()` to assign mapped reads to genomic features

First, we will install the **RSubread** package. For the sake of knitting the document, the package installs are commented out

```

#Install "RSubread" package
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("Rsubread")

```

```
library(Rsubread)
```

```
## Warning: package 'Rsubread' was built under R version 4.0.2
```

We will be counting the reads that are mapped to the orphan TSSs of *S. venezuelae* for our previously sorted .bam files. There are three time points per strain (corresponding to vegetative, fragmentation, and sporulation


```

## ||   Min overlapping bases : 1                               ||
## ||                                           ||
## ||           Chimeric reads : counted                       ||
## ||           Both ends mapped : not required                ||
## ||                                           ||
## \\=====\\
##
## //===== Running =====\\
## ||
## || Load annotation file 10h_orphan.gff ...                ||
## ||   Features : 313                                         ||
## ||   Meta-features : 313                                    ||
## ||   Chromosomes/contigs : 1                                ||
## ||
## || Process BAM file rnc_10h_A.bam...                         ||
## ||   Paired-end reads are included.                         ||
## ||   Total alignments : 1095070                             ||
## ||   Successfully assigned alignments : 53533 (4.9%)         ||
## ||   Running time : 0.08 minutes                            ||
## ||
## || Process BAM file rnc_10h_B.bam...                         ||
## ||   Paired-end reads are included.                         ||
## ||   Total alignments : 989847                               ||
## ||   Successfully assigned alignments : 50609 (5.1%)         ||
## ||   Running time : 0.07 minutes                            ||
## ||
## || Process BAM file WT_10h_A.bam...                         ||
## ||   Paired-end reads are included.                         ||
## ||   Total alignments : 985345                               ||
## ||   Successfully assigned alignments : 42802 (4.3%)         ||
## ||   Running time : 0.07 minutes                            ||
## ||
## || Process BAM file WT_10h_B.bam...                         ||
## ||   Paired-end reads are included.                         ||
## ||   Total alignments : 1038457                             ||
## ||   Successfully assigned alignments : 41261 (4.0%)         ||
## ||   Running time : 0.07 minutes                            ||
## ||
## || Write the final count table.                             ||
## || Write the read assignment summary.                       ||
## ||
## \\=====\\

```

```
veg_10h_rnc_counts <- data.frame(veg_10h_rnc$counts)
```

vegetative vs 14h TSSs

```

setwd("filter_bam/rnc/veg/")

veg_14h_rnc <- featureCounts(files=dir(".", "bam$"),
                             annot.ext= "~/Desktop/RNA_seq/ref/14h_orphan.gff",
                             isGTFAnnotationFile=TRUE,
                             GTF.featureType="orphan_TSS",
                             GTF.attrType="ID",

```

```
useMetaFeatures = FALSE,  
isPairedEnd = TRUE)
```

```

##          =====
##          ===== /-----\-----\-----\-----\-----\-----\
##          ===== | (-----|-----|-----|-----|-----|-----|
##          ===== \-----\-----\-----\-----\-----\-----\
##          ===== )-----|-----|-----|-----\-----\-----\
##          ===== |-----/-----/-----/-----\-----\-----/
## Rsubread 2.2.6
##
## //===== featureCounts setting =====\
## ||
## ||          Input files : 4 BAM files
## ||                      o rnc_10h_A.bam
## ||                      o rnc_10h_B.bam
## ||                      o WT_10h_A.bam
## ||                      o WT_10h_B.bam
## ||
## ||          Annotation : 14h_orphan.gff (GTF)
## ||          Dir for temp files : .
## ||          Threads : 1
## ||          Level : feature level
## ||          Paired-end : yes
## ||          Multimapping reads : counted
## ||          Multi-overlapping reads : not counted
## ||          Min overlapping bases : 1
## ||
## ||          Chimeric reads : counted
## ||          Both ends mapped : not required
## ||
## \=====
##
## //===== Running =====\
## ||
## ||          Load annotation file 14h_orphan.gff ...
## ||          Features : 387
## ||          Meta-features : 387
## ||          Chromosomes/contigs : 1
## ||
## ||          Process BAM file rnc_10h_A.bam...
## ||          Paired-end reads are included.
## ||          Total alignments : 1095070
## ||          Successfully assigned alignments : 23612 (2.2%)
## ||          Running time : 0.07 minutes
## ||
## ||          Process BAM file rnc_10h_B.bam...
## ||          Paired-end reads are included.
## ||          Total alignments : 989847
## ||          Successfully assigned alignments : 21664 (2.2%)
## ||          Running time : 0.06 minutes
## ||
## ||          Process BAM file WT_10h_A.bam...

```



```

## ||      Multimapping reads : counted      ||
## || Multi-overlapping reads : not counted  ||
## ||      Min overlapping bases : 1         ||
## ||                                         ||
## ||      Chimeric reads : counted          ||
## ||      Both ends mapped : not required   ||
## ||                                         ||
## \\=====//
##
## //===== Running =====\\
## ||
## || Load annotation file 14h_orphan.gff ...
## ||      Features : 387
## ||      Meta-features : 387
## ||      Chromosomes/contigs : 1
## ||
## || Process BAM file rnc_14h_A.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 870222
## ||      Successfully assigned alignments : 35867 (4.1%)
## ||      Running time : 0.12 minutes
## ||
## || Process BAM file rnc_14h_B.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 866786
## ||      Successfully assigned alignments : 22426 (2.6%)
## ||      Running time : 0.13 minutes
## ||
## || Process BAM file WT_14h_A.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 690612
## ||      Successfully assigned alignments : 12643 (1.8%)
## ||      Running time : 0.05 minutes
## ||
## || Process BAM file WT_14h_B.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 691538
## ||      Successfully assigned alignments : 14484 (2.1%)
## ||      Running time : 0.05 minutes
## ||
## || Write the final count table.
## || Write the read assignment summary.
## ||
## \\=====//

```

```
frag_14h_rnc_counts <- data.frame(frag_14h_rnc$counts)
```

fragmentation vs 18h TSSs

```

setwd("filter_bam/rnc/frag/")

frag_18h_rnc <- featureCounts(files=dir(".", "bam$"),
                             annot.ext= "~/Desktop/RNA_seq/ref/18h_orphan.gff",
                             isGTFAnnotationFile=TRUE,

```

```
## =====  
## ===== /-----| | | | \----- \----- \----- \-----  
## ===== | (-----| | | | )-----| | | | /-----| | | |  
## ===== \-----| | | | <| | | /-----| | | | /-----| | | |  
## ===== )-----| | | | )-----| | | | \-----| | | |  
## ===== |-----/ \-----/|-----/| | | \-----/ /----- \-----/  
## Rsubread 2.2.6  
##  
## //===== featureCounts setting =====\\  
## ||  
## || Input files : 4 BAM files ||  
## || o rnc_14h_A.bam ||  
## || o rnc_14h_B.bam ||  
## || o WT_14h_A.bam ||  
## || o WT_14h_B.bam ||  
## ||  
## || Annotation : 18h_orphan.gff (GTF) ||  
## || Dir for temp files : . ||  
## || Threads : 1 ||  
## || Level : feature level ||  
## || Paired-end : yes ||  
## || Multimapping reads : counted ||  
## || Multi-overlapping reads : not counted ||  
## || Min overlapping bases : 1 ||  
## ||  
## || Chimeric reads : counted ||  
## || Both ends mapped : not required ||  
## ||  
## \\=====\\  
##  
## //===== Running =====\\  
## ||  
## || Load annotation file 18h_orphan.gff ... ||  
## || Features : 381 ||  
## || Meta-features : 381 ||  
## || Chromosomes/contigs : 1 ||  
## ||  
## || Process BAM file rnc_14h_A.bam... ||  
## || Paired-end reads are included. ||  
## || Total alignments : 870222 ||  
## || Successfully assigned alignments : 33006 (3.8%) ||  
## || Running time : 0.10 minutes ||  
## ||  
## || Process BAM file rnc_14h_B.bam... ||  
## || Paired-end reads are included. ||  
## || Total alignments : 866786 ||  
## || Successfully assigned alignments : 30406 (3.5%) ||  
## || Running time : 0.10 minutes ||
```



```

## || Multi-overlapping reads : not counted ||
## ||   Min overlapping bases : 1 ||
## || ||
## \\=====\\
##
## //===== Running =====\\
## ||
## || Load annotation file 24h_orphan.gff ... ||
## ||   Features : 410 ||
## ||   Meta-features : 410 ||
## ||   Chromosomes/contigs : 1 ||
## || ||
## || Process BAM file rnc_20h_A.bam... ||
## ||   Single-end reads are included. ||
## ||   Total alignments : 1045402 ||
## ||   Successfully assigned alignments : 39617 (3.8%) ||
## ||   Running time : 0.02 minutes ||
## || ||
## || Process BAM file rnc_20h_B.bam... ||
## ||   Single-end reads are included. ||
## ||   Total alignments : 925635 ||
## ||   Successfully assigned alignments : 38175 (4.1%) ||
## ||   Running time : 0.02 minutes ||
## || ||
## || Process BAM file WT_20h_A.bam... ||
## ||   Single-end reads are included. ||
## ||   Total alignments : 10730651 ||
## ||   Successfully assigned alignments : 272691 (2.5%) ||
## ||   Running time : 0.23 minutes ||
## || ||
## || Process BAM file WT_20h_B.bam... ||
## ||   Single-end reads are included. ||
## ||   Total alignments : 8851854 ||
## ||   Successfully assigned alignments : 243852 (2.8%) ||
## ||   Running time : 0.20 minutes ||
## || ||
## || Write the final count table. ||
## || Write the read assignment summary. ||
## || ||
## \\=====\\
spore_24h_rnc_counts <- data.frame(spore_24h_rnc$counts)

```

DESeq2: are these orphan TSSs deferentially expressed?

Next, we must prepare the information corresponding to the columns of the count matrix or the sample information (colData). To do this we will create a data frame with the names of each .bam file in one column and their corresponding genotype, “mutant” or “wildtype”, in the second column.

veg_10h_rnc file prep

```

# Create a vector containing the names of the .bam files (will be column one)
files_veg_10h_rnc <- veg_10h_rnc$targets

```

```
# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_veg_10h_rnc <- c("mutant1","mutant1","wildtype", "wildtype")
```

```
# Create the data frame "coldata"
coldata_veg_10h_rnc <- data.frame(files_veg_10h_rnc, genotype_veg_10h_rnc)
```

veg_14h_rnc file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_veg_14h_rnc <- veg_14h_rnc$targets
```

```
# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_veg_14h_rnc <- c("mutant1","mutant1", "wildtype", "wildtype")
```

```
# Create the data frame "coldata"
coldata_veg_14h_rnc <- data.frame(files_veg_14h_rnc, genotype_veg_14h_rnc)
```

frag_14h_rnc file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_frag_14h_rnc <- frag_14h_rnc$targets
```

```
# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_frag_14h_rnc <- c("mutant1","mutant1","wildtype", "wildtype")
```

```
# Create the data frame "coldata"
coldata_frag_14h_rnc <- data.frame(files_frag_14h_rnc, genotype_frag_14h_rnc)
```

frag_18h_rnc file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_frag_18h_rnc <- frag_18h_rnc$targets
```

```
# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_frag_18h_rnc <- c("mutant1","mutant1","wildtype", "wildtype")
```

```
# Create the data frame "coldata"
coldata_frag_18h_rnc <- data.frame(files_frag_18h_rnc, genotype_frag_18h_rnc)
```

spore_24h_rnc file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_spore_24h_rnc <- spore_24h_rnc$targets
```

```
# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_spore_24h_rnc <- c("mutant1","mutant1","wildtype", "wildtype")
```

```
# Create the data frame "coldata"
coldata_spore_24h_rnc <- data.frame(files_spore_24h_rnc, genotype_spore_24h_rnc)
```

Now that we have prepared the input for the countData and colData arguments of the DESeqDataSetFromMatrix() function, we can generate the DESeqDataSet. Specify counts as the input for countData, coldata as the input for colData and ~genotype as the input for design, as we are interested in comparing differentially expressed genes between the mutant and wildtype.

```
library(DESeq2)
```

```

# Generate the DESeqDataSet, "deseqdata"
#for comparing differentially expressed genes in the mutant and wildtype

#veg_10h_rnc
deseqdata_veg_10h_rnc <- DESeqDataSetFromMatrix(countData = veg_10h_rnc_counts,
                                                colData = coldata_veg_10h_rnc,
                                                design = ~genotype_veg_10h_rnc)

#veg_14h
deseqdata_veg_14h_rnc <- DESeqDataSetFromMatrix(countData = veg_14h_rnc_counts,
                                                colData = coldata_veg_14h_rnc,
                                                design = ~genotype_veg_14h_rnc)

#frag_14h
deseqdata_frag_14h_rnc <- DESeqDataSetFromMatrix(countData = frag_14h_rnc_counts,
                                                colData = coldata_frag_14h_rnc,
                                                design = ~genotype_frag_14h_rnc)

#frag_18h
deseqdata_frag_18h_rnc <- DESeqDataSetFromMatrix(countData = frag_18h_rnc_counts,
                                                colData = coldata_frag_18h_rnc,
                                                design = ~genotype_frag_18h_rnc)

#frag_24h
deseqdata_spore_24h_rnc <- DESeqDataSetFromMatrix(countData = spore_24h_rnc_counts,
                                                colData = coldata_spore_24h_rnc,
                                                design = ~genotype_spore_24h_rnc)

```

Differential expression analysis!

We will be using the standard `DeSeq()` function and associated functions to perform differential expression analysis. Since the last variable in our design formula above is a factor, we must specify that the “wildtype” should be used to as the reference level for analysis. By default, `DeSeq2` will choose the reference level based on whichever factor comes first in the alphabet. To assign the “wildtype” as the reference level, re-level the factors using the `factor()` function.

```

# Relevel the factors in "genotype" so that wildtype is the reference level for analysis
deseqdata_veg_10h_rnc$genotype_veg_10h_rnc <- factor(deseqdata_veg_10h_rnc$genotype_veg_10h_rnc,
                                                    levels = c("wildtype", "mutant1"))

deseqdata_veg_14h_rnc$genotype_veg_14h_rnc <- factor(deseqdata_veg_14h_rnc$genotype_veg_14h_rnc,
                                                    levels = c("wildtype", "mutant1"))

deseqdata_frag_14h_rnc$genotype_frag_14h_rnc <- factor(deseqdata_frag_14h_rnc$genotype_frag_14h_rnc,
                                                    levels = c("wildtype", "mutant1"))

deseqdata_frag_18h_rnc$genotype_frag_18h_rnc <- factor(deseqdata_frag_18h_rnc$genotype_frag_18h_rnc,
                                                    levels = c("wildtype", "mutant1"))

deseqdata_spore_24h_rnc$genotype_spore_24h_rnc <- factor(deseqdata_spore_24h_rnc$genotype_spore_24h_rnc,
                                                    levels = c("wildtype", "mutant1"))

```

Next, we will pass the `DESeqDataSet` to `DeSeq2` for differential gene analysis.

DESeq function

```
# Perform DESeq() analysis and output to results table
deseqdata_veg_10h_rnc <- DESeq(object = deseqdata_veg_10h_rnc)
deseqdata_veg_14h_rnc <- DESeq(object = deseqdata_veg_14h_rnc)
deseqdata_frag_14h_rnc <- DESeq(object = deseqdata_frag_14h_rnc)
deseqdata_frag_18h_rnc <- DESeq(object = deseqdata_frag_18h_rnc)
deseqdata_spore_24h_rnc <- DESeq(object = deseqdata_spore_24h_rnc)

# Generate the results table
results_veg_10h_rnc <- results(deseqdata_veg_10h_rnc)
results_veg_14h_rnc <- results(deseqdata_veg_14h_rnc)
results_frag_14h_rnc <- results(deseqdata_frag_14h_rnc)
results_frag_18h_rnc <- results(deseqdata_frag_18h_rnc)
results_spore_24h_rnc <- results(deseqdata_spore_24h_rnc)

#turn the table into a data frame
veg_10h_rnc_dataframe <- data.frame(results_veg_10h_rnc)
veg_14h_rnc_dataframe <- data.frame(results_veg_14h_rnc)
frag_14h_rnc_dataframe <- data.frame(results_frag_14h_rnc)
frag_18h_rnc_dataframe <- data.frame(results_frag_18h_rnc)
spore_24h_rnc_dataframe <- data.frame(results_spore_24h_rnc)
```

We have results!!!

The next order of business will be to subset these 5 tables by their adj-pvalue and logFC

subsetting DESeq2 results table

```
#filter by p-value less than 0.05 and logFC > 2
library(dplyr)

veg_10h_rnc_dataframe <- filter(veg_10h_rnc_dataframe,
                                padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
veg_14h_rnc_dataframe <- filter(veg_14h_rnc_dataframe,
                                padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
frag_14h_rnc_dataframe <- filter(frag_14h_rnc_dataframe,
                                padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
frag_18h_rnc_dataframe <- filter(frag_18h_rnc_dataframe,
                                padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
spore_24h_rnc_dataframe <- filter(spore_24h_rnc_dataframe,
                                padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)

#filter out NA's
veg_10h_rnc_dataframe <- na.omit(veg_10h_rnc_dataframe)
veg_14h_rnc_dataframe <- na.omit(veg_14h_rnc_dataframe)
frag_14h_rnc_dataframe <- na.omit(frag_14h_rnc_dataframe)
frag_18h_rnc_dataframe <- na.omit(frag_18h_rnc_dataframe)
spore_24h_rnc_dataframe <- na.omit(spore_24h_rnc_dataframe)
```

Woohoo!! The last step for the RNase III mutant is to write csv files, which we can use for further analysis down the road!

Write csvs for RNase III mutant

```
write.csv(veg_10h_rnc_dataframe,
          file = "~/Desktop/RNA_seq/DESeq2/RNaseIII_veg_10hTSS.csv",
          quote = FALSE)

write.csv(veg_14h_rnc_dataframe,
          file = "~/Desktop/RNA_seq/DESeq2/RNaseIII_veg_14hTSS.csv",
          quote = FALSE)

write.csv(frag_14h_rnc_dataframe,
          file = "~/Desktop/RNA_seq/DESeq2/RNaseIII_frag_14hTSS.csv",
          quote = FALSE)

write.csv(frag_18h_rnc_dataframe,
          file = "~/Desktop/RNA_seq/DESeq2/RNaseIII_frag_18hTSS.csv",
          quote = FALSE)

write.csv(spore_24h_rnc_dataframe,
          file = "~/Desktop/RNA_seq/DESeq2/RNaseIII_spore_24hTSS.csv",
          quote = FALSE)
```

Next up: the exact same thing but with RNase J

vegetative vs 10h TSSs

```
setwd("filter_bam/rnj/veg/")

veg_10h_rnj <- featureCounts(files=dir(".", "bam$"),
                             annot.ext= "~/Desktop/RNA_seq/ref/10h_orphan.gff",
                             isGTFAnnotationFile=TRUE,
                             GTF.featureType="orphan_TSS",
                             GTF.attrType="ID",
                             useMetaFeatures = FALSE,
                             isPairedEnd = TRUE)

##
## =====
## ===== /-----| | | |-----\-----|-----/ \-----\
## ===== | (-----| | | | |) | |) | |-----/ \-----| | | |
## ===== \-----| | | | |<| |-----/ |-----/ \-----| | | |
## ===== ) | | | | |) | | \ | |-----/-----\ | | | |
## ===== |-----/ \-----/ |-----/ | | \ \-----/ / \ \-----/
## Rsubread 2.2.6
##
## //===== featureCounts setting =====\\
## ||
## ||           Input files : 4 BAM files
## ||               o rnj_12h_A.bam
## ||               o rnj_12h_B.bam
## ||               o WT_10h_A.bam
## ||               o WT_10h_B.bam
## ||
## ||           Annotation : 10h_orphan.gff (GTF)
## ||
```

```

## ||      Dir for temp files : .
## ||      Threads : 1
## ||      Level : feature level
## ||      Paired-end : yes
## ||      Multimapping reads : counted
## ||      Multi-overlapping reads : not counted
## ||      Min overlapping bases : 1
## ||
## ||      Chimeric reads : counted
## ||      Both ends mapped : not required
## ||
## \\=====//
##
## //===== Running =====\\
## ||
## || Load annotation file 10h_orphan.gff ...
## ||      Features : 313
## ||      Meta-features : 313
## ||      Chromosomes/contigs : 1
## ||
## || Process BAM file rnj_12h_A.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 909054
## ||      Successfully assigned alignments : 54427 (6.0%)
## ||      Running time : 0.06 minutes
## ||
## || Process BAM file rnj_12h_B.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 641196
## ||      Successfully assigned alignments : 38113 (5.9%)
## ||      Running time : 0.04 minutes
## ||
## || Process BAM file WT_10h_A.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 985345
## ||      Successfully assigned alignments : 42802 (4.3%)
## ||      Running time : 0.06 minutes
## ||
## || Process BAM file WT_10h_B.bam...
## ||      Paired-end reads are included.
## ||      Total alignments : 1038457
## ||      Successfully assigned alignments : 41261 (4.0%)
## ||      Running time : 0.07 minutes
## ||
## || Write the final count table.
## || Write the read assignment summary.
## ||
## \\=====//
veg_10h_rnj_counts <- data.frame(veg_10h_rnj$counts)

```

vegetative vs 14h TSSs

```
setwd("filter_bam/rnj/veg/")

veg_14h_rnj <- featureCounts(files=dir(".", "bam$"),
                             annot.ext= "~/Desktop/RNA_seq/ref/14h_orphan.gff",
                             isGTFAnnotationFile=TRUE,
                             GTF.featureType="orphan_TSS",
                             GTF.attrType="ID",
                             useMetaFeatures = FALSE,
                             isPairedEnd = TRUE)
```

```
##
##
## =====
## ===== /-----| | | | \-----|-----| \-----| \-----|
## ===== | (-----| | | | |_) | |_) | |_) | |_) | |_) | |_) |
## ===== \-----| | | | |<| | | | | | | | | | | | | | | | | |
## ===== | |_) | |_) | |_) | |_) | |_) | |_) | |_) | |_) | |_) |
## ===== |-----/ \-----/ |-----/ | | \ \-----/ / \ \-----/
## Rsubread 2.2.6
##
## //===== featureCounts setting =====\\
## ||
## ||          Input files : 4 BAM files
## ||                      o rnj_12h_A.bam
## ||                      o rnj_12h_B.bam
## ||                      o WT_10h_A.bam
## ||                      o WT_10h_B.bam
## ||
## ||          Annotation : 14h_orphan.gff (GTF)
## ||          Dir for temp files : .
## ||          Threads : 1
## ||          Level : feature level
## ||          Paired-end : yes
## ||          Multimapping reads : counted
## ||          Multi-overlapping reads : not counted
## ||          Min overlapping bases : 1
## ||
## ||          Chimeric reads : counted
## ||          Both ends mapped : not required
## ||
## \\=====\\
##
## //===== Running =====\\
## ||
## || Load annotation file 14h_orphan.gff ...
## ||   Features : 387
## ||   Meta-features : 387
## ||   Chromosomes/contigs : 1
## ||
## || Process BAM file rnj_12h_A.bam...
## ||   Paired-end reads are included.
## ||   Total alignments : 909054
## ||   Successfully assigned alignments : 17259 (1.9%)
## ||   Running time : 0.06 minutes
## ||
```

```
## || Process BAM file rnj_12h_B.bam... ||
## || Paired-end reads are included. ||
## || Total alignments : 641196 ||
## || Successfully assigned alignments : 12168 (1.9%) ||
## || Running time : 0.04 minutes ||
## ||
## || Process BAM file WT_10h_A.bam... ||
## || Paired-end reads are included. ||
## || Total alignments : 985345 ||
## || Successfully assigned alignments : 20394 (2.1%) ||
## || Running time : 0.06 minutes ||
## ||
## || Process BAM file WT_10h_B.bam... ||
## || Paired-end reads are included. ||
## || Total alignments : 1038457 ||
## || Successfully assigned alignments : 20394 (2.0%) ||
## || Running time : 0.07 minutes ||
## ||
## || Write the final count table. ||
## || Write the read assignment summary. ||
## ||
## \\=====//
```

#now extract the counts table from the output!

```
veg_14h_rnj_counts <- data.frame(veg_14h_rnj$counts)
```

fragmentation vs 14h TSSs

```
setwd("filter_bam/rnj/frag/")
```

```
frag_14h_rnj <- featureCounts(files=dir(".", "bam$"),
                              annot.ext= "~/Desktop/RNA_seq/ref/14h_orphan.gff",
                              isGTFAnnotationFile=TRUE,
                              GTF.featureType="orphan_TSS",
                              GTF.attrType="ID",
                              useMetaFeatures = FALSE,
                              isPairedEnd = TRUE)
```

```
##
## =====
## ===== / _ _ _ | | | _ _ _ \ _ _ _ \ _ _ _ \ _ _ _ \
## ===== | ( _ _ | | | | ) | | ) | | _ / \ | | | |
## ===== \ _ _ \ | | | | _ < | _ / | | | / \ \ | | | |
## ===== _ _ _ ) | | _ | | ) | | \ \ | | _ _ _ / _ _ _ \ | | | |
## ===== | _ _ _ / \ _ _ _ / _ _ _ / | | \ \ _ _ _ _ / _ _ _ \ \ _ _ _ /
## Rsubread 2.2.6
##
## //===== featureCounts setting =====\\
## ||
## || Input files : 4 BAM files ||
## || o rnj_18h_A.bam ||
## || o rnj_18h_B.bam ||
## || o WT_14h_A.bam ||
```

```

## ||                               o WT_14h_B.bam                               ||
## ||                               ||                                           ||
## ||               Annotation : 14h_orphan.gff (GTF)                         ||
## ||       Dir for temp files : .                                           ||
## ||               Threads : 1                                             ||
## ||               Level : feature level                                     ||
## ||               Paired-end : yes                                         ||
## ||       Multimapping reads : counted                                     ||
## ||       Multi-overlapping reads : not counted                           ||
## ||       Min overlapping bases : 1                                         ||
## ||                               ||                                           ||
## ||               Chimeric reads : counted                                 ||
## ||               Both ends mapped : not required                         ||
## ||                               ||                                           ||
## ||=====\\
##
## //===== Running =====\\
## ||
## || Load annotation file 14h_orphan.gff ...
## ||       Features : 387
## ||       Meta-features : 387
## ||       Chromosomes/contigs : 1
## ||
## || Process BAM file rnj_18h_A.bam...
## ||       Paired-end reads are included.
## ||       Total alignments : 887135
## ||       Successfully assigned alignments : 9833 (1.1%)
## ||       Running time : 0.06 minutes
## ||
## || Process BAM file rnj_18h_B.bam...
## ||       Paired-end reads are included.
## ||       Total alignments : 992954
## ||       Successfully assigned alignments : 26157 (2.6%)
## ||       Running time : 0.11 minutes
## ||
## || Process BAM file WT_14h_A.bam...
## ||       Paired-end reads are included.
## ||       Total alignments : 690612
## ||       Successfully assigned alignments : 12643 (1.8%)
## ||       Running time : 0.05 minutes
## ||
## || Process BAM file WT_14h_B.bam...
## ||       Paired-end reads are included.
## ||       Total alignments : 691538
## ||       Successfully assigned alignments : 14484 (2.1%)
## ||       Running time : 0.05 minutes
## ||
## || Write the final count table.
## || Write the read assignment summary.
## ||
## ||=====\\
frag_14h_rnj_counts <- data.frame(frag_14h_rnj$counts)

```

```
setwd("filter_bam/rnj/frag/")

frag_18h_rnj <- featureCounts(files=dir(".", "bam$"),
                              annot.ext= "~/Desktop/RNA_seq/ref/18h_orphan.gff",
                              isGTFAnnotationFile=TRUE,
                              GTF.featureType="orphan_TSS",
                              GTF.attrType="ID",
                              useMetaFeatures = FALSE,
                              isPairedEnd = TRUE)
```

30


```

## ||          Annotation : 24h_orphan.gff (GTF) ||
## ||      Dir for temp files : . ||
## ||          Threads : 1 ||
## ||          Level : feature level ||
## ||      Paired-end : no ||
## ||      Multimapping reads : counted ||
## || Multi-overlapping reads : not counted ||
## ||      Min overlapping bases : 1 ||
## || ||
## ||=====\\
##
## //===== Running =====\\
## ||
## || Load annotation file 24h_orphan.gff ... ||
## ||      Features : 410 ||
## ||      Meta-features : 410 ||
## ||      Chromosomes/contigs : 1 ||
## || ||
## || Process BAM file rnj_24h_A.bam... ||
## ||      Single-end reads are included. ||
## ||      Total alignments : 1071922 ||
## ||      Successfully assigned alignments : 37032 (3.5%) ||
## ||      Running time : 0.02 minutes ||
## || ||
## || Process BAM file rnj_24h_B.bam... ||
## ||      Single-end reads are included. ||
## ||      Total alignments : 1261624 ||
## ||      Successfully assigned alignments : 40935 (3.2%) ||
## ||      Running time : 0.02 minutes ||
## || ||
## || Process BAM file WT_20h_A.bam... ||
## ||      Single-end reads are included. ||
## ||      Total alignments : 10730651 ||
## ||      Successfully assigned alignments : 272691 (2.5%) ||
## ||      Running time : 0.23 minutes ||
## || ||
## || Process BAM file WT_20h_B.bam... ||
## ||      Single-end reads are included. ||
## ||      Total alignments : 8851854 ||
## ||      Successfully assigned alignments : 243852 (2.8%) ||
## ||      Running time : 0.19 minutes ||
## || ||
## || Write the final count table. ||
## || Write the read assignment summary. ||
## || ||
## ||=====\\
spore_24h_rnj_counts <- data.frame(spore_24h_rnj$counts)

```

DESeq2: are these orphan TSSs deferentially expressed?

Next, we must prepare the information corresponding to the columns of the count matrix or the sample information (colData). To do this we will create a data frame with the names of each .bam file in one column

and their corresponding genotype, “mutant” or “wildtype”, in the second column.

veg_10h_rnj file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_veg_10h_rnj <- veg_10h_rnj$targets

# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_veg_10h_rnj <- c("mutant2","mutant2","wildtype", "wildtype")

# Create the data frame "coldata"
coldata_veg_10h_rnj <- data.frame(files_veg_10h_rnj, genotype_veg_10h_rnj)
```

veg_14h_rnj file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_veg_14h_rnj <- veg_14h_rnj$targets

# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_veg_14h_rnj <- c("mutant2","mutant2", "wildtype", "wildtype")

# Create the data frame "coldata"
coldata_veg_14h_rnj <- data.frame(files_veg_14h_rnj, genotype_veg_14h_rnj)
```

frag_14h_rnj file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_frag_14h_rnj <- frag_14h_rnj$targets

# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_frag_14h_rnj <- c("mutant2","mutant2","wildtype", "wildtype")

# Create the data frame "coldata"
coldata_frag_14h_rnj <- data.frame(files_frag_14h_rnj, genotype_frag_14h_rnj)
```

frag_18h_rnj file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_frag_18h_rnj <- frag_18h_rnj$targets

# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_frag_18h_rnj <- c("mutant2","mutant2","wildtype", "wildtype")

# Create the data frame "coldata"
coldata_frag_18h_rnj <- data.frame(files_frag_18h_rnj, genotype_frag_18h_rnj)
```

spore_24h_rnj file prep

```
# Create a vector containing the names of the .bam files (will be column one)
files_spore_24h_rnj <- spore_24h_rnj$targets

# Create a vector containing the corresponding genotypes of the above .bam files (will be column two)
genotype_spore_24h_rnj <- c("mutant2","mutant2","wildtype", "wildtype")

# Create the data frame "coldata"
coldata_spore_24h_rnj <- data.frame(files_spore_24h_rnj, genotype_spore_24h_rnj)
```

Generate a DESeq DataSet

Now that we have prepared the input for the countData and colData arguments of the DESeqDataSetFromMatrix() function, we can generate the DESeqDataSet. Specify counts as the input for countData, coldata as the input for colData and ~genotype as the input for design, as we are interested in comparing differentially expressed genes between the mutant and wildtype.

```
library(DESeq2)

# Generate the DESeqDataSet, "deseqdata"
# for comparing differentially expressed genes in the mutant and wildtype

# veg_10h_rnj
deseqdata_veg_10h_rnj <- DESeqDataSetFromMatrix(countData = veg_10h_rnj_counts,
                                                colData = coldata_veg_10h_rnj,
                                                design = ~genotype_veg_10h_rnj)

# veg_14h
deseqdata_veg_14h_rnj <- DESeqDataSetFromMatrix(countData = veg_14h_rnj_counts,
                                                colData = coldata_veg_14h_rnj,
                                                design = ~genotype_veg_14h_rnj)

# frag_14h
deseqdata_frag_14h_rnj <- DESeqDataSetFromMatrix(countData = frag_14h_rnj_counts,
                                                colData = coldata_frag_14h_rnj,
                                                design = ~genotype_frag_14h_rnj)

# frag_18h
deseqdata_frag_18h_rnj <- DESeqDataSetFromMatrix(countData = frag_18h_rnj_counts,
                                                colData = coldata_frag_18h_rnj,
                                                design = ~genotype_frag_18h_rnj)

# frag_24h
deseqdata_spore_24h_rnj <- DESeqDataSetFromMatrix(countData = spore_24h_rnj_counts,
                                                colData = coldata_spore_24h_rnj,
                                                design = ~genotype_spore_24h_rnj)
```

Differential expression analysis!

We will be using the standard DESeq() function and associated functions to perform differential expression analysis. Since the last variable in our design formula above is a factor, we must specify that the “wildtype” should be used to as the reference level for analysis. By default, DESeq2 will choose the reference level based on whichever factor comes first in the alphabet. To assign the “wildtype” as the reference level, re-level the factors using the factor() function.

```
# Relevel the factors in "genotype" so that wildtype is the reference level for analysis
deseqdata_veg_10h_rnj$genotype_veg_10h_rnj <- factor(deseqdata_veg_10h_rnj$genotype_veg_10h_rnj,
                                                    levels = c("wildtype", "mutant2"))

deseqdata_veg_14h_rnj$genotype_veg_14h_rnj <- factor(deseqdata_veg_14h_rnj$genotype_veg_14h_rnj,
                                                    levels = c("wildtype", "mutant2"))

deseqdata_frag_14h_rnj$genotype_frag_14h_rnj <- factor(deseqdata_frag_14h_rnj$genotype_frag_14h_rnj,
                                                    levels = c("wildtype", "mutant2"))

deseqdata_frag_18h_rnj$genotype_frag_18h_rnj <- factor(deseqdata_frag_18h_rnj$genotype_frag_18h_rnj,
                                                    levels = c("wildtype", "mutant2"))
```

```
deseqdata_spore_24h_rnj$genotype_spore_24h_rnj <- factor(deseqdata_spore_24h_rnj$genotype_spore_24h_rnj
  levels = c("wildtype", "mutant2"))
```

Next, we will pass the DESeqDataSet to DESeq2 for differential gene analysis.

DESeq function

```
# Perform DESeq() analysis and output to results table
deseqdata_veg_10h_rnj <- DESeq(object = deseqdata_veg_10h_rnj)
deseqdata_veg_14h_rnj <- DESeq(object = deseqdata_veg_14h_rnj)
deseqdata_frag_14h_rnj <- DESeq(object = deseqdata_frag_14h_rnj)
deseqdata_frag_18h_rnj <- DESeq(object = deseqdata_frag_18h_rnj)
deseqdata_spore_24h_rnj <- DESeq(object = deseqdata_spore_24h_rnj)

# Generate the results table
results_veg_10h_rnj <- results(deseqdata_veg_10h_rnj)
results_veg_14h_rnj <- results(deseqdata_veg_14h_rnj)
results_frag_14h_rnj <- results(deseqdata_frag_14h_rnj)
results_frag_18h_rnj <- results(deseqdata_frag_18h_rnj)
results_spore_24h_rnj <- results(deseqdata_spore_24h_rnj)

#turn the table into a data frame
veg_10h_rnj_dataframe <- data.frame(results_veg_10h_rnj)
veg_14h_rnj_dataframe <- data.frame(results_veg_14h_rnj)
frag_14h_rnj_dataframe <- data.frame(results_frag_14h_rnj)
frag_18h_rnj_dataframe <- data.frame(results_frag_18h_rnj)
spore_24h_rnj_dataframe <- data.frame(results_spore_24h_rnj)
```

We have results!!!

The next order of business will be to subset these 5 tables by their adj-pvalue and logFC

subsetting DESeq2 results table

```
#filter by adj p-value<0.05 and logFC>2 and baseMean>50
library(dplyr)
veg_10h_rnj_dataframe <- filter(veg_10h_rnj_dataframe,
  padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
veg_14h_rnj_dataframe <- filter(veg_14h_rnj_dataframe,
  padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
frag_14h_rnj_dataframe <- filter(frag_14h_rnj_dataframe,
  padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
frag_18h_rnj_dataframe <- filter(frag_18h_rnj_dataframe,
  padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)
spore_24h_rnj_dataframe <- filter(spore_24h_rnj_dataframe,
  padj<0.05 & abs(log2FoldChange)>2 & baseMean>50)

#filter out NA's
veg_10h_rnj_dataframe <- na.omit(veg_10h_rnj_dataframe)
veg_14h_rnj_dataframe <- na.omit(veg_14h_rnj_dataframe)
frag_14h_rnj_dataframe <- na.omit(frag_14h_rnj_dataframe)
```

```
frag_18h_rnj_dataframe <- na.omit(frag_18h_rnj_dataframe)
spore_24h_rnj_dataframe <- na.omit(spore_24h_rnj_dataframe)
```

Woo! The last step for the RNase J mutant is to write csv files, which we can use for further analysis down the road!

Write csvs for RNase J mutant

```
write.csv(veg_10h_rnj_dataframe,
  file = "~/Desktop/RNA_seq/DESeq2/RNaseJ_veg_10hTSS.csv",
  quote = FALSE)

write.csv(veg_14h_rnj_dataframe,
  file = "~/Desktop/RNA_seq/DESeq2/RNaseJ_veg_14hTSS.csv",
  quote = FALSE)

write.csv(frag_14h_rnj_dataframe,
  file = "~/Desktop/RNA_seq/DESeq2/RNaseJ_frag_14hTSS.csv",
  quote = FALSE)

write.csv(frag_18h_rnj_dataframe,
  file = "~/Desktop/RNA_seq/DESeq2/RNaseJ_frag_18hTSS.csv",
  quote = FALSE)

write.csv(spore_24h_rnj_dataframe,
  file = "~/Desktop/RNA_seq/DESeq2/RNaseJ_spore_24hTSS.csv",
  quote = FALSE)
```

That's it for this markdown! To look at differential expression of actual annotated *S. venezuelae* genes we will be reading in DESeq2 outputs that were generated previously by Emma M (MSc graduate in the Elliot lab)

For the next and final section of this cumulative project, please see the “integration” markdown!