

PRJNA297798

Guillermo Ayala

2024-04-02

Table of contents

Data	1
Download the sra files	1
Generating the fastq files	2
Download the index file for bowtie2	3
Generating the directories with the index files	3
Short read alignment using bowtie2	3
From sam to bam using samtools and sorting the reads	4
Using Rsamtools for counting the reads aligned.	5
Creating files with the names of the bam files.	5
Executing R code	5

Data

This data has been analyzed in

>Kaletsky, R.; Lakhina, V.; Arey, R.; Williams, A.; Landis, J.; Ashraf, J. & Murphy, C. T.
The C. elegans adult neuronal IIS/FOXO transcriptome reveals adult phenotype regulators
Nature, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights
Reserved., 2016, 529, 92-96

Download the sra files

```
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143370
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143371
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143371
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143372
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143372
```

```
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143373
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143373
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143383
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143383
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143386
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143386
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143386
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143386
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143387
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByExp/sra/SRX/SRX143/SRX143388
```

Generating the fastq files

```
fastq-dump -I --split-files SRR2969230.sra
fastq-dump -I --split-files SRR2969231.sra
fastq-dump -I --split-files SRR2969232.sra
fastq-dump -I --split-files SRR2969233.sra
fastq-dump -I --split-files SRR2969234.sra
fastq-dump -I --split-files SRR2969235.sra
fastq-dump -I --split-files SRR2969236.sra
fastq-dump -I --split-files SRR2969237.sra
fastq-dump -I --split-files SRR2969238.sra
fastq-dump -I --split-files SRR2969239.sra
fastq-dump -I --split-files SRR2969240.sra
fastq-dump -I --split-files SRR2969241.sra
fastq-dump -I --split-files SRR2969242.sra
fastq-dump -I --split-files SRR2969243.sra
fastq-dump -I --split-files SRR2969244.sra
fastq-dump -I --split-files SRR2969245.sra
fastq-dump -I --split-files SRR2969246.sra
fastq-dump -I --split-files SRR2979473.sra
fastq-dump -I --split-files SRR2979475.sra
fastq-dump -I --split-files SRR2979476.sra
```

Download the index file for bowtie2

```
wget ftp://igenome:G3nom3s4u@uscd-ftp.illumina.com/Caenorhabditis_elegans/Ensembl/WBcel235
```

Generating the directories with the index files

```
gzip -d Caenorhabditis_elegans_Ensembl_WBcel235.tar.gz  
tar xvf Caenorhabditis_elegans_Ensembl_WBcel235.tar.gz
```

Short read alignment using bowtie2

```
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -1 SRR2969  
rm SRR2969230_*.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -1 SRR2969  
rm SRR2969231_*.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -1 SRR2969  
rm SRR2969232_*.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969233_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969234_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969235_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -1 SRR2969  
rm SRR2969236_*.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -1 SRR2969  
rm SRR2969237_*.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -1 SRR2969  
rm SRR2969238_*.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969239_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969240_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969241_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969  
rm SRR2969242_1.fastq  
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969
```

```

rm SRR2969243_1.fastq
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969
rm SRR2969244_1.fastq
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969
rm SRR2969245_1.fastq
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2969
rm SRR2969246_1.fastq
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2979
rm SRR2979473_1.fastq
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2979
rm SRR2979475_1.fastq
bowtie2 -x Caenorhabditis_elegans/Ensembl/WBcel235/Sequence/Bowtie2Index/genome -U SRR2979
rm SRR2979476_1.fastq

```

From sam to bam using samtools and sorting the reads

```

samtools view -bS SRR2969230.sam | samtools sort - SRR2969230
rm SRR2969230.sam
samtools view -bS SRR2969231.sam | samtools sort - SRR2969231
rm SRR2969231.sam
samtools view -bS SRR2969232.sam | samtools sort - SRR2969232
rm SRR2969232.sam
samtools view -bS SRR2969233.sam | samtools sort - SRR2969233
rm SRR2969233.sam
samtools view -bS SRR2969234.sam | samtools sort - SRR2969234
rm SRR2969234.sam
samtools view -bS SRR2969235.sam | samtools sort - SRR2969235
rm SRR2969235.sam
samtools view -bS SRR2969236.sam | samtools sort - SRR2969236
rm SRR2969236.sam
samtools view -bS SRR2969237.sam | samtools sort - SRR2969237
rm SRR2969237.sam
samtools view -bS SRR2969238.sam | samtools sort - SRR2969238
rm SRR2969238.sam
samtools view -bS SRR2969239.sam | samtools sort - SRR2969239
rm SRR2969239.sam
samtools view -bS SRR2969240.sam | samtools sort - SRR2969240
rm SRR2969240.sam
samtools view -bS SRR2969241.sam | samtools sort - SRR2969241
rm SRR2969241.sam

```

```

samtools view -bS SRR2969242.sam | samtools sort - SRR2969242
rm SRR2969242.sam
samtools view -bS SRR2969243.sam | samtools sort - SRR2969243
rm SRR2969243.sam
samtools view -bS SRR2969244.sam | samtools sort - SRR2969244
rm SRR2969244.sam
samtools view -bS SRR2969245.sam | samtools sort - SRR2969245
rm SRR2969245.sam
samtools view -bS SRR2969246.sam | samtools sort - SRR2969246
rm SRR2969246.sam
samtools view -bS SRR2979473.sam | samtools sort - SRR2979473
rm SRR2979473.sam
samtools view -bS SRR2979475.sam | samtools sort - SRR2979475
rm SRR2979475.sam
samtools view -bS SRR2979476.sam | samtools sort - SRR2979476
rm SRR2979476.sam

```

Using Rsamtools for counting the reads aligned.

Creating files with the names of the bam files.

Some samples are single and other samples are pair-ended. We create the file BamFiles.txt with the following content.

```

SRR2969246.bam SRR2969245.bam SRR2969244.bam SRR2969243.bam SRR2969242.bam
SRR2969241.bam SRR2969240.bam SRR2969239.bam SRR2969235.bam SRR2979476.bam
SRR2969234.bam SRR2979475.bam SRR2969233.bam SRR2979473.bam SRR2969233.bam
SRR2979473.bam SRR2969238.bam SRR2969237.bam SRR2969236.bam SRR2969232.bam
SRR2969231.bam SRR2969230.bam

```

Executing R code

```

library(Rsamtools)
library(GenomicFeatures)
library(GenomicAlignments)
gtfFile = "Caenorhabditis_elegans/Ensembl/WBcel235/Annotation/Genes/genes.gtf"
txdb = makeTxDbFromGFF(gtfFile, format="gtf")
genes = exonsBy(txdb, by="gene")
dirActualData = paste(getwd(), "/", sep="")
sampleTableSingle = read.table("BamFiles.txt")

```

```

fls = paste(dirActualData,sampleTableSingle[,1],sep="")
bamLst = BamFileList(fl, index=character(),yieldSize=100000,obeyQname=TRUE)
PRJNA297798 = summarizeOverlaps(features = genes,read=bamLst,
  mode="Union",
  singleEnd=TRUE,
  ignore.strand=TRUE,
  fragments=FALSE)
metadatos = read.csv("PRJNA297798coldata.csv",header=TRUE,sep=",")
metadatos[, "strain"] = factor(metadatos[, "strain"],levels=1:4,
  labels=c("daf-2(e1370);daf-16(mu86);Punc-119::gfp",
    "daf-2(e1370);Punc-119::gfp",
    "pmec-4::gfp","Punc-119::gfp"))
metadatos[, "tissue"] = factor(metadatos[, "tissue"],levels=1:2,labels=c("Neurons","Whole
metadatos[, "treatment"] = factor(metadatos[, "treatment"],levels=1:6,
  labels=c("daf-2(e1370); daf-16(mu86); Punc-119::gfp--sorted neurons",
    "daf-2(e1370) neurons","N2; Pmec-4::GFP--sorted mechanosensory neurons"
    "daf-2(e1370); Punc-119::gfp--sorted neurons","N2; Punc-119::gfp--sorte
    "N2; Punc-119::gfp--unsorted/whole worm"))
colData(PRJNA297798) = DataFrame(metadatos)
colnames(PRJNA297798) = colData(PRJNA297798)[, "run"]
save(PRJNA297798,file="PRJNA297798.rda")

```

We have a `SummarizedExperiment::RangedSummarizedExperiment` object.