

PRJNA266927

Guillermo Ayala

2025-03-11

Table of contents

Download the fastq files from ENA	1
Download the index file for bowtie and bowtie2	1
Generating the directories with the index files	2
Short read alignment using bowtie2	2
From sam to bam using samtools and sorting the reads	2
Creating files with the names of the bam files.	3
Using Rsamtools for counting the reads aligned.	3
Short read alignment using samtools	3

Download the fastq files from ENA

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/005/SRR1656855/SRR1656855.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/006/SRR1656856/SRR1656856.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/007/SRR1656857/SRR1656857.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/008/SRR1656858/SRR1656858.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/009/SRR1656859/SRR1656859.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/000/SRR1656860/SRR1656860.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/001/SRR1656861/SRR1656861.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR165/002/SRR1656862/SRR1656862.fastq.gz
```

Download the index file for bowtie and bowtie2

```
wget ftp://igenome:G3nom3s4u@ussd-ftp.illumina.com/Drosophila_melanogaster/Ensembl/BDGP6/D
```

Generating the directories with the index files

```
gzip -d Drosophila_melanogaster_Ensembl_BDGP6.tar.gz
tar xvf Drosophila_melanogaster_Ensembl_BDGP6.tar
```

Short read alignment using bowtie2

```
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656855
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656856
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656857
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656858
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656859
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656860
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656861
bowtie2 -x Drosophila_melanogaster/Ensembl/BDGP6/Sequence/Bowtie2Index/genome -U SRR1656862
```

From sam to bam using samtools and sorting the reads

```
samtools view -bS SRR1656855.sam | samtools sort - SRR1656855
rm SRR1656855.sam
samtools view -bS SRR1656856.sam | samtools sort - SRR1656856
rm SRR1656856.sam
samtools view -bS SRR1656857.sam | samtools sort - SRR1656857
rm SRR1656857.sam
samtools view -bS SRR1656858.sam | samtools sort - SRR1656858
rm SRR1656858.sam
samtools view -bS SRR1656859.sam | samtools sort - SRR1656859
rm SRR1656859.sam
samtools view -bS SRR1656860.sam | samtools sort - SRR1656860
rm SRR1656860.sam
samtools view -bS SRR1656861.sam | samtools sort - SRR1656861
rm SRR1656861.sam
samtools view -bS SRR1656862.sam | samtools sort - SRR1656862
rm SRR1656862.sam
```

Creating files with the names of the bam files.

SRR1656855.bam
SRR1656856.bam
SRR1656857.bam
SRR1656858.bam
SRR1656859.bam
SRR1656860.bam
SRR1656861.bam
SRR1656862.bam

Using Rsamtools for counting the reads aligned.

```
library(Rsamtools)
library(GenomicFeatures)
library(GenomicAlignments)
gtfFile = "Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf"
txdb = makeTxDbFromGFF(gtfFile, format="gtf")
genes = exonsBy(txdb, by="gene")
dirActualData = paste(getwd(), "/", sep="")
sampleTableSingle = read.table("BamSingle.txt")
fls = paste(dirActualData, sampleTableSingle[,1], sep="")
bamLst = BamFileList(fls, index=character(), yieldSize=100000, obeyQname=TRUE)
PRJNA266927 = summarizeOverlaps(features = genes, read=bamLst,
  mode="Union",
  singleEnd=TRUE,
  ignore.strand=TRUE,
  fragments=FALSE)
Run = c("SRR1656855", "SRR1656856", "SRR1656857", "SRR1656858", "SRR1656859", "SRR1656860",
  "SRR1656861", "SRR1656862")
Treatment = c(1,1,2,2,3,3,4,4)
Treatment = factor(Treatment, levels=1:4, labels=c("NC", "NOTCH", "ESG", "TUMOR"))
colData(PRJNA266927) = DataFrame(Treatment)
save(PRJNA266927, file="PRJNA266927.rda")
```

Short read alignment using samtools

What else can we obtain using samtools?

```
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
tophat -p 8 -G Drosophila_melanogaster/Ensembl/BDGP6/Annotation/Genes/genes.gtf -o SRR1656
```