# Commando’s voor STAR (herkansing)

Door Moshtach

In de annotations map staan de homo sapiens data en annotatie data.

## Unzippen van annotatie en humane genoom

gunzip -k gencode.v29.annotation\_chr10.gtf.gz

gunzip -k Homo\_sapiens.GRCh38.dna.chromosome.10.fa.gz

## Index maken

Cd

mkdir indexes

mkdir indexes/chr10

$RUN STAR --runMode genomeGenerate --genomeDir indexes/chr10 --genomeFastaFiles annotations/Homo\_sapiens.GRCh38.dna.chromosome.10.fa --sjdbGTFfile annotations/gencode.v29.annotation\_chr10.gtf --sjdbOverhang 50 --outFileNamePrefix chr10

## Aligning reads voor iedere sample

mkdir alignments

### A549\_0\_1

$RUN STAR --genomeDir indexes/chr10 --readFilesIn resources/A549\_0\_1chr10\_1.fastq.gz resources/A549\_0\_1chr10\_2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --quantMode GeneCounts --outFileNamePrefix alignments/A549\_0\_1

### A549\_0\_2

$RUN STAR --genomeDir indexes/chr10 --readFilesIn resources/A549\_0\_2chr10\_1.fastq.gz resources/A549\_0\_2chr10\_2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --quantMode GeneCounts --outFileNamePrefix alignments/A549\_0\_2

### A549\_0\_3

$RUN STAR --genomeDir indexes/chr10 --readFilesIn resources/A549\_0\_3chr10\_1.fastq.gz resources/A549\_0\_3chr10\_2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --quantMode GeneCounts --outFileNamePrefix alignments/A549\_0\_3

### A549\_25\_1

$RUN STAR --genomeDir indexes/chr10 --readFilesIn resources/A549\_25\_1chr10\_1.fastq.gz resources/A549\_25\_1chr10\_2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --quantMode GeneCounts --outFileNamePrefix alignments/A549\_25\_1

### A549\_25\_2

$RUN STAR --genomeDir indexes/chr10 --readFilesIn resources/A549\_25\_2chr10\_1.fastq.gz resources/A549\_25\_2chr10\_2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --quantMode GeneCounts --outFileNamePrefix alignments/A549\_25\_2

### A549\_25\_3

$RUN STAR --genomeDir indexes/chr10 --readFilesIn resources/A549\_25\_3chr10\_1.fastq.gz resources/A549\_25\_3chr10\_2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --quantMode GeneCounts --outFileNamePrefix alignments/A549\_25\_3

### Counts voor iedere sample:

De counts zitten in een .tab file in de alignments mapje.

Outputfiles bekijken:

more alignments/A549\_0\_1ReadsPerGene.out.tab

more alignments/A549\_0\_2ReadsPerGene.out.tab

more alignments/A549\_0\_3ReadsPerGene.out.tab

more alignments/A549\_25\_1ReadsPerGene.out.tab

more alignments/A549\_25\_2ReadsPerGene.out.tab

more alignments/A549\_25\_3ReadsPerGene.out.tab

## Matrix

Om van alle outputfiles een grote bestand te maken zijn de volgende commando’s uitgevoerd:

mkdir full\_data2

mkdir deseq2 (in de directory v full\_data2)

mkdir counts\_star (in de dir v full\_data2)

**~/alignments**$ paste A549\_\*ReadsPerGene.out.tab | grep -v "\_" | awk '{printf "%s\t", $1}{for (i=4;i<=NF;i+=4) printf "%s\t", $i; printf "\n" }' > tmp

**~/alignments**$ sed -e "1igene\_name\t$(ls A549\_\*ReadsPerGene.out.tab | tr '\n' '\t' | sed 's/ReadsPerGene.out.tab//g')" ~/alignments/tmp | cut -f1-7 > ~/full\_data2/deseq2/raw\_counts\_A549\_matrix.txt

Output bestand bekijken: cat raw\_counts\_A549\_matrix.txt

Deze bestand is vervolgens genormaliseerd in de Python script: normalized\_counts.py