STAR/hisat2 mapping

Samtools Sam to Bam

Htseq/FeatureCounts/Stringtie/TEtranscripts read counting

DEseq2/ballgown analysis

FastQC/MultiQC QC

**Alignment**

Create genome index files:

nohup STAR --runMode genomeGenerate --genomeDir ./genome --genomeFastaFiles ~/data/Drosophila\_melanogaster.BDGP6.32.dna.toplevel.fa --sjdbGTFfile ~/data/Drosophila\_melanogaster.BDGP6.32.106.gtf &

Alignment:

nohup STAR --genomeDir ~/data/learn/1/genome \

--readFilesIn DfM4\_1.fq.gz DfM4\_2.fq.gz \

--outFileNamePrefix ~/data/learn/1/alignment/DfM4 \

--outSAMtype BAM SortedByCoordinate \

--readFilesCommand zcat \

--runThreadN 20 \

--limitBAMsortRAM 7563196624 \

--winAnchorMultimapNmax 200 \

--outFilterMultimapNmax 100 & (the last two for TEtranscripts)

Bam index:

nohup samtools index DfM4Aligned.sortedByCoord.out.bam &

**Counting**

nohup htseq-count -f bam -r pos -s no ~/data/learn/1/alignment/DfM4Aligned.sortedByCoord.out.bam ~/data/dm6.ncbiRefSeq.gtf > DfM4output\_file.txt &

TEtranscripts

nohup TEtranscripts --sortByPos --format BAM --mode multi \

-t ~/data/learn/2/alignment/DfMB1Aligned.sortedByCoord.out.bam \

~/data/learn/2/alignment/DfMB2Aligned.sortedByCoord.out.bam \

~/data/learn/2/alignment/DfMB3Aligned.sortedByCoord.out.bam \

~/data/learn/2/alignment/DfMB4Aligned.sortedByCoord.out.bam \

-c ~/data/learn/2/alignment/CSWB1Aligned.sortedByCoord.out.bam \

~/data/learn/2/alignment/CSWB2Aligned.sortedByCoord.out.bam \

~/data/learn/2/alignment/CSWB3Aligned.sortedByCoord.out.bam \

~/data/learn/2/alignment/CSWB4Aligned.sortedByCoord.out.bam \

--GTF ~/data/dm6.ncbiRefSeq.gtf \

--TE ~/data/dm6\_rmsk\_TE.gtf &

**Analysis**

DEseq2

Enhanced volcano

**GO**

EGO:Biological Process/Molecular Functions/Cellular Components

KEGG

GSEA