# install fastqc

bioinfmsc5:~$ wget <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.8.zip>

#unzip the zip file of fastqc

bioinfmsc5:~$ unzip fastqc\_v0.11.8.zip

#change the permission of fastqc

bioinfmsc5:~$ chmod ug+x FastQC/fastqc

#find the remote fastq files and copy to my own directory

bioinfmsc5:~$ cd /localdisk/data/BPSM/Assignment1/fastq

bioinfmsc5:~$ mkdir zyy

bioinfmsc5:~$ cp/localdisk/data/BPSM/Assignment1/fastq/216\_L8\_1.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/216\_L8\_2.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/218\_L8\_1.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/218\_L8\_2.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/219\_L8\_1.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/219\_L8\_2.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/220\_L8\_1.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/220\_L8\_2.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/221\_L8\_1.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/222\_L8\_1.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/221\_L8\_2.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/222\_L8\_2.fq.gz zyy

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/fastq/fqfiles zyy

bioinfmsc5:~$ mkdir fastqc\_output

#use fastqc to access the quality of the raw sequences

bioinfmsc5:~$ FastQC/fastqc zyy/216\_L8\_1.fq.gz -o fastqc\_output

bioinfmsc5:~$ FastQC/fastqc zyy/216\_L8\_2.fq.gz -o fastqc\_output

# use for loop to do the rest quality tests for 218-222 files

bioinfmsc5:~$ for i in `echo 218 219 220 221 222`

> do

> echo quality control $i...

> FastQC/fastqc zyy/${i}\_L8\_1.fq.gz -o fastqc\_output

> FastQC/fastqc zyy/${i}\_L8\_2.fq.gz -o fastqc\_output

> done

# for the bowtie2, it needs to bulid a reference index to read, so first is to build a index

bioinfmsc5:~$ mkdir referenceindex

bioinfmsc5:~$ cp /localdisk/data/BPSM/Assignment1/Tbb\_genome/ referenceindex

bioinfmsc5:~$ cd referenceindex

# for bowtie2, the zipped files cannot be read, it needs to be fasta format, so unzip it.

bioinfmsc5:~/referenceindex$ gunzip Tb927\_genome.fasta.gz

bioinfmsc5:~/referenceindex$ cd

bioinfmsc5:~$ bowtie2-build referenceindex/Tb927\_genome.fasta Trypanosoma\_brucei

bioinfmsc5:~$ mv Trypanosoma\_brucei.1.bt2 referenceindex

bioinfmsc5:~$ mv Trypanosoma\_brucei.2.bt2 referenceindex

bioinfmsc5:~$ mv Trypanosoma\_brucei.3.bt2 referenceindex

bioinfmsc5:~$ mv Trypanosoma\_brucei.4.bt2 referenceindex

bioinfmsc5:~$ mv Trypanosoma\_brucei.rev.1.bt2 referenceindex

bioinfmsc5:~$ mv Trypanosoma\_brucei.rev.2.bt2 referenceindex

# use 8 threads to speed up the rate, tried the originally rate, was too slow….

# use bowtie2 command to map the alignments, and then output.

# for each sample, repeat the same step because I failed to write a for loop command

bioinfmsc5:~$ bowtie2 –p 8 –x referenceindex/Trypanosoma\_brucei -1 zyy/216\_L8\_1.fq.gz -2 zyy/216\_L8\_2.fq.gz -S 216\_L8\_output.sam

bioinfmsc5:~$ mkdir samfiles

bioinfmsc5:~$ mv 216\_L8\_output.sam samfiles

bioinfmsc5:~$ bowtie2 -p 8 -x referenceindex/Trypanosoma\_brucei -1 zyy/218\_L8\_1.fq.gz -2 zyy/218\_L8\_2.fq.gz -S 218\_L8\_output.sam

bioinfmsc5:~$ mv 218\_L8\_output.sam samfiles

bioinfmsc5:~$ bowtie2 -p 8 -x referenceindex/Trypanosoma\_brucei -1 zyy/219\_L8\_1.fq.gz -2 zyy/219\_L8\_2.fq.gz -S 219\_L8\_output.sam

bioinfmsc5:~$ mv 219\_L8\_output.sam samfiles

bioinfmsc5:~$ bowtie2 -p 8 -x referenceindex/Trypanosoma\_brucei -1 zyy/220\_L8\_1.fq.gz -2 zyy/220\_L8\_2.fq.gz -S 220\_L8\_output.sam

bioinfmsc5:~$ mv 220\_L8\_output.sam samfiles

bioinfmsc5:~$ bowtie2 -p 8 -x referenceindex/Trypanosoma\_brucei -1 zyy/221\_L8\_1.fq.gz -2 zyy/221\_L8\_2.fq.gz -S 221\_L8\_output.sam

bioinfmsc5:~$ mv 221\_L8\_output.sam samfiles

bioinfmsc5:~$ bowtie2 -p 8 -x referenceindex/Trypanosoma\_brucei -1 zyy/222\_L8\_1.fq.gz -2 zyy/222\_L8\_2.fq.gz -S 222\_L8\_output.sam

bioinfmsc5:~$ mv 222\_L8\_output.sam samfiles

# for the bedtools, it need the bam format input, so transfer the sam to bam

# before use the command, files need to be sorted

# use bedtools genomecov to ensure the outcome if the alignment is more than once

# note: bedtools intersect match for only one aligenment, so I’m not very sure

# then output to bedGraph files, do for each sample

bioinfmsc5:~$ mv tbb\_bed/Tbbgenes.bed samfiles

bioinfmsc5:~$ cd samfiles

bioinfmsc5:~/samfiles$ samtools view -bS 216\_L8\_output.sam>216\_L8\_output.bam

bioinfmsc5:~/samfiles$ samtools sort 216\_L8\_output.sam > 216\_L8\_output.sam.sorted.bam

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 216\_L8\_output.sam.sorted.bam -g Tbbgenes.bed

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 216\_L8\_output.sam.sorted.bam -g Tbbgenes.bed>216\_L8.bedGraph

bioinfmsc5:~/samfiles$ samtools view -bS 216\_L8\_output.sam>218\_L8\_output.bam

bioinfmsc5:~/samfiles$ samtools sort 216\_L8\_output.sam > 218\_L8\_output.sam.sorted.bam

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 218\_L8\_output.sam.sorted.bam -g Tbbgenes.bed

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 218\_L8\_output.sam.sorted.bam -g Tbbgenes.bed>216\_L8.bedGraph

bioinfmsc5:~/samfiles$ samtools view -bS 219\_L8\_output.sam>219\_L8\_output.bam

bioinfmsc5:~/samfiles$ samtools sort 219\_L8\_output.sam > 219\_L8\_output.sam.sorted.bam

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 219\_L8\_output.sam.sorted.bam -g Tbbgenes.bed

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 219\_L8\_output.sam.sorted.bam -g Tbbgenes.bed>219\_L8.bedGraph

bioinfmsc5:~/samfiles$ samtools view -bS 220\_L8\_output.sam>220\_L8\_output.bam

bioinfmsc5:~/samfiles$ samtools sort 220\_L8\_output.sam > 220\_L8\_output.sam.sorted.bam

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 220\_L8\_output.sam.sorted.bam -g Tbbgenes.bed

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 220\_L8\_output.sam.sorted.bam -g Tbbgenes.bed>220\_L8.bedGraph

bioinfmsc5:~/samfiles$ samtools view -bS 221\_L8\_output.sam>221\_L8\_output.bam

bioinfmsc5:~/samfiles$ samtools sort 221\_L8\_output.sam > 221\_L8\_output.sam.sorted.bam

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 221\_L8\_output.sam.sorted.bam -g Tbbgenes.bed

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 221\_L8\_output.sam.sorted.bam -g Tbbgenes.bed>221\_L8.bedGraph

bioinfmsc5:~/samfiles$ samtools view -bS 222\_L8\_output.sam>222\_L8\_output.bam

bioinfmsc5:~/samfiles$ samtools sort 222\_L8\_output.sam > 222\_L8\_output.sam.sorted.bam

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 222\_L8\_output.sam.sorted.bam -g Tbbgenes.bed

bioinfmsc5:~/samfiles$ bedtools genomecov -bg -ibam 222\_L8\_output.sam.sorted.bam -g Tbbgenes.bed>222\_L8.bedGraph

# calculate the mean value of each alignment, the result show at the last column

# do for each sample

bioinfmsc5:~$ cd samfiles

bioinfmsc5:~/samfiles$ bedtools map -a Tbbgenes.bed -b 216\_L8.bedGraph -c 4 -o mean

bioinfmsc5:~/samfiles$ bedtools map -a Tbbgenes.bed -b 218\_L8.bedGraph -c 4 -o mean

bioinfmsc5:~/samfiles$ bedtools map -a Tbbgenes.bed -b 219\_L8.bedGraph -c 4 -o mean

bioinfmsc5:~/samfiles$ bedtools map -a Tbbgenes.bed -b 220\_L8.bedGraph -c 4 -o mean

bioinfmsc5:~/samfiles$ bedtools map -a Tbbgenes.bed -b 221\_L8.bedGraph -c 4 -o mean

bioinfmsc5:~/samfiles$ bedtools map -a Tbbgenes.bed -b 222\_L8.bedGraph -c 4 -o mean