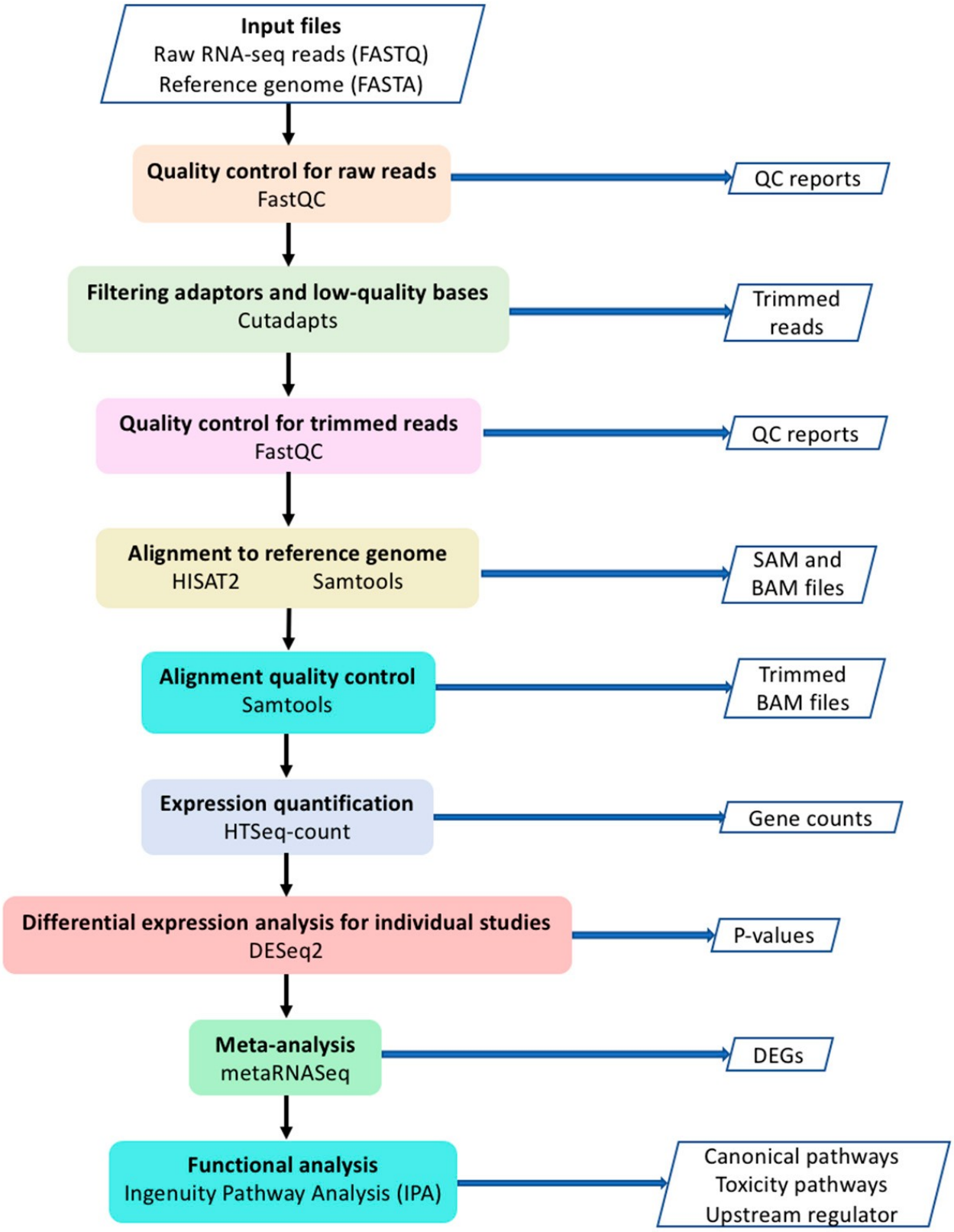
RNA_SEQ WORKFLOW



RNA_SEQ DATA_ANALYSIS USING CAMMANDS =

RNA_SEQ

work only rna-seq
ENA === bisulfite-seq plants
(Study View all 61 results.
Study Accession no. == SRP072503(click this acession no.)Project: PRJNA316697
Bisulfite-Seq and RNA-seq profiling of Arabidopsis aneuploids)
Organism: Arabidopsis thaliana
Secondary Study Accession: SRP072503
Study Title:
Bisulfite-Seq and RNA-seq profiling of Arabidopsis aneuploids
Total read file = 53(bisulfite-seq and rna-seq)
open sra == SRP072503 (paste read accession no.)
source-rnaseq (21)- send to - file -format - accession no.(then download total accession no. rana-sq)
install enabrowser tool

tolal rna-seq reads=
SRR3309279
SRR3309280
SRR3309281
SRR3309299
SRR3309300
SRR3309301
SRR3309302
SRR3309303
SRR3309304
SRR3309305
SRR3309306
SRR3309307
SRR3309308
SRR3309309
SRR3309310
SRR3309311
SRR3309312
SRR3309313
SRR3309314
SRR3309315
SRR3309316

1) FastQC (quality control of raw reads)

cat rr |while read i; do cd $i; fastqc *.gz ; cd .. ; done(create html file) for multiple files work
create files==SRR3309310_1_fastqc.html SRR3309310_1_fastqc.zip SRR3309310_1.fastq.gz SRR3309310_2_fastqc.html SRR3309310_2_fastqc.zip SRR3309310_2.fastq.gz)
for single file work =====> fastqc *.gz

2) Trimmomatic or trim_galore (filtering adapter and low quality bases)
for paired end=====>
according to===> java -jar trimmomatic-0.39.jar PE input_forward.fq.gz input_reverse.fq.gz output_forward_paired.fq.gz output_forward_unpaired.fq.gz output_reverse_paired.fq.gz output_reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 MINLEN:36
for one reads files ==>
java -jar trimmomatic-0.36.jar PE SRR3309279_1.fastq.gz SRR3309279_2.fastq.gz 79_forward_paired.fq.gz 79_forward_unpaired.fq.gz 79_reverse_paired.fq.gz 79_reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2 LEADING:3 TRAILING:3 MINLEN:36

for multiple read files ==========>
cat rr | while read i ; do cd $i ; java -jar trimmomatic-0.36.jar PE *1.fastq.gz *2.fastq.gz forward_paired.fq.gz forward_unpaired.fq.gz reverse_paired.fq.gz reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2 LEADING:3 TRAILING:3 MINLEN:20 ; cd .. ; done(error)
cat rr | while read i ; do cp TruSeq3-PE.fa $i ; done (all read files in copy TruSeq3-PE.fa)
cat rr | while read i ; do cd $i ; java -jar ../trimmomatic-0.36.jar PE *1.fastq.gz *2.fastq.gz forward_paired.fq.gz forward_unpaired.fq.gz reverse_paired.fq.gz reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2 LEADING:3 TRAILING:3 MINLEN:20 ; cd .. ; done
trimmomatic files run in fastqc==========>
cat rr | while read i ; do cd $i ; fastqc *1.fastq.gz *2.fastq.gz forward_paired.fq.gz forward_unpaired.fq.gz reverse_paired.fq.gz reverse_unpaired.fq.gz ; cd.. ; done
cat rr | while read i ; do cd $i ; fastqc forward_paired.fq.gz forward_unpaired.fq.gz reverse_paired.fq.gz reverse_unpaired.fq.gz ; cd.. ; done
cat rr | while read i ; do cd $i ; fastqc *.gz ; cd .. ; done(ALL .GZ FILE)

2) Trim_galore

USAGE: trim_galore [options] <filename(s)>
option== --stringency 13 -q 30 --paired , --gzip
trim_galore --stringency 13 -q 30 --paired --gzip SRR3309279_1.fastq.gz SRR3309279_2.fastq.gz -o trim_galore(single reads)
cat rr | while read i; do cd $i ; trim_galore --stringency 13 -q 30 --paired *1.fastq.gz *2.fastq.gz -o trim_galore_$i ;cd .. ; done(all reads)
trim file run in fastq for check adapter remove or not
catr rr |while read i; do cd $i ; fastqc *.gz ; cd .. ; done(create html files of raw reads(quality control for trimimng reads))

3) bowtie2(alignment to referenc genome)

The `bowtie2-build` indexer

The bowtie2 build indexer==========
first download reference genome of Arabidopsis thaliana.
NCBI - datbase - genome - search txid3702[orgn] (Arabidopsis thaliana = GCF_000001735.4_TAIR10.1_genomic.fna.gz)
##bowtie2-build [options]* <reference_in> <bt2_base> (option= -fThe reference input files (specified as <reference_in>) are FASTA files (usually having extension .fa, .mfa, .fna or similar).
bowtie2-build [options]* <reference_in> <bt2_base>
cammand==========>
bowtie2-build -f GCF_000001735.4_TAIR10.1_genomic.fna.gz bt2_base

Aligning reads

bowtie2 -x bt2_base -1 SRR3309279_2_val_2.fq.gz -2 SRR3309279_1_val_1.fq.gz -S trim_galore_SRR3309279.sam

cat list |while read i ; do cd $i ; bowtie2 -x ../../Bowtie/index/bt2_base -1 *1.fq.gz -2 *2.fq.gz -S $i\.sam; cd .. ; done

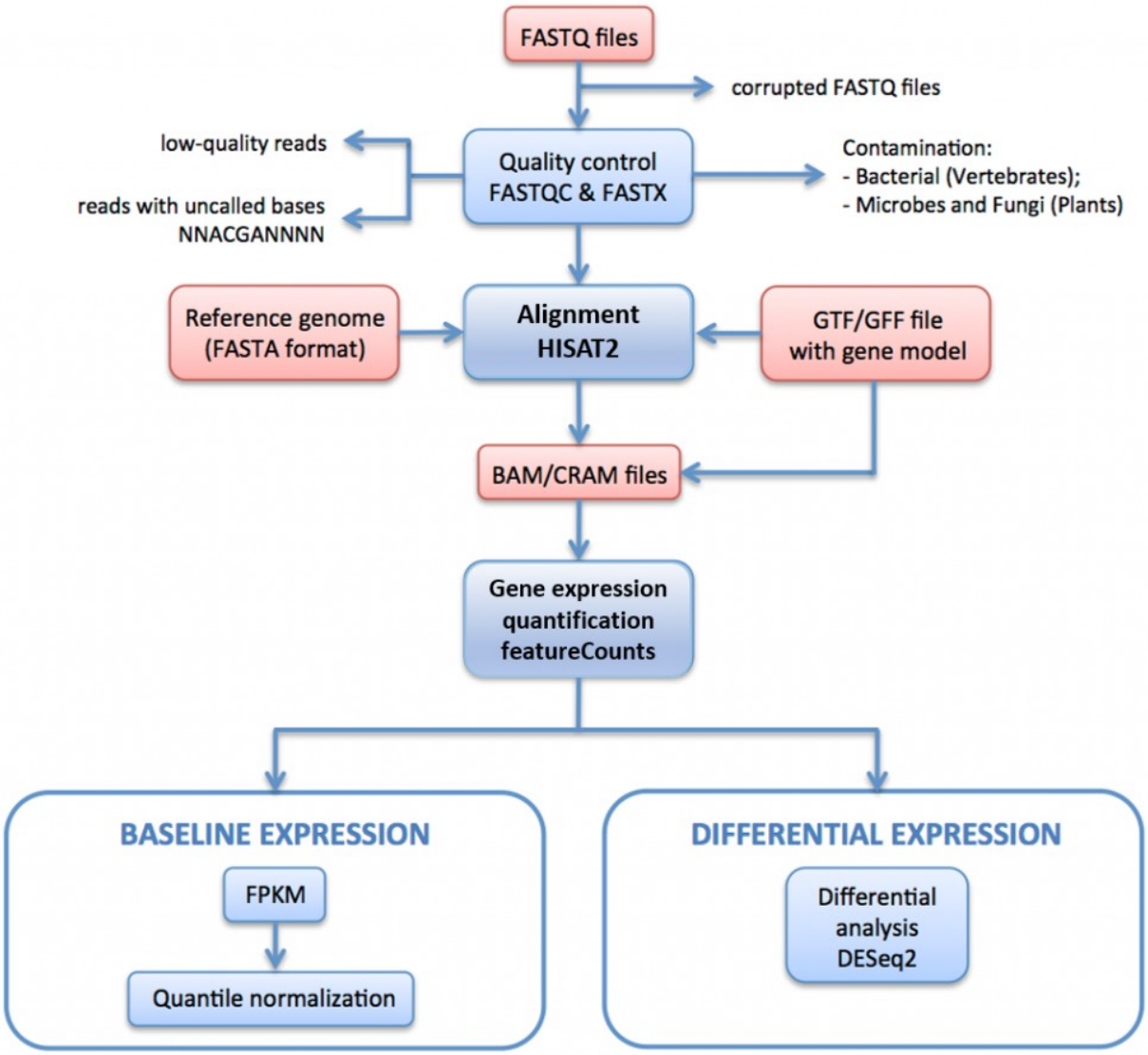
create = .sam file
list = total accesion files
reference genome path = ../../Bowtie/index/bt2_base
cat list |sed "1,12d" |while read i ; do cd $i ; bowtie2 -x ../../Bowtie/index/bt2_base -1 *1.fq.gz -2 *2.fq.gz -p 20 S $i.sam; cd .. ; done (use processor)

cat list |while read i ; do cd $i ; bowtie2 -x ../../Bowtie/index/bt2_base -1 *1.fq.gz -2 *2.fq.gz -p 20 -S $i.sam; cd .. ; done

4) Samtools

samtools view -S -b sample.sam > sample.bam

samtools sort sample.bam -o sample.sorted.bam(sort bam file)
#Now let's check the order:
samtools view sample.sorted.bam | head
samtools view sample.sorted.bam | wc -l(#count total alignment= 10076007)
samtools view -H sample.sorted.bam(inspect header)
samtools view forward62.sorted.bam | awk '{print $5}' | sort -k1g| uniq -c
samtools view forward62.sorted.bam | awk '{print $2}' | sort | uniq -c
samtools view -f 0*2 forward62.sorted.bam |head(capture the FLAG)
samtools view -f 0*2 forward62.sorted.bam |wc -l(how mony properly alignment there)


FASTQ files
corrupted FASTQ files
low-quality reads
reads with uncalled bases
NNACGANNNN
Quality control
FASTQC & FASTX
Contamination:
- Bacterial (Vertebrates);
- Microbes and Fungi (Plants)
Reference genome
(FASTA format)
Alignment
HISAT2
GTF/GFF file
with gene model
BAM/CRAM files
Gene expression
quantification
featureCounts
BASELINE EXPRESSION
FPKM
Quantile normalization
DIFFERENTIAL EXPRESSION
Differential
analysis
DESeq2