

RNA-seq 순서

■ download samples and reference genome files

Ensembl에서 다운 받은 자료는 read 데이터

■ Trimming (Trimmomatic)

Fragmentation된 fragment들을 paired end read한 데이터에서 양 끝 어댑터 혹은 퀄리티가 낮은 데이터를 제거하기 위한 단계

■ Quality control (FastQC)

Trim된 read의 퀄리티를 확인하기 위한 단계

■ Indexing

순서 상관없이 분해된 fragment = read 들을 mapping하기 전에 한 read와 동일한 염기서열을 reference genome에서 일일이 찾는 것보다 인덱싱해둔 reference genome을 참고했을 때 더 빠르게 찾을 수 있기 때문에 인덱싱을 만들어 두는 단계(사전처럼)

■ Alignment (HISAT2, STAR)

Reference genome에 각 reads 들을 매핑하는 단계

■ Sam to Bam to sorted Bam(samtools)

Mapping을 통해 얻은 sam 파일을 정렬한 뒤, 용량이 너무 크기 때문에 binary 형식의 bam파일로 변환하는 단계

■ Assembly (stringtie)

bam 파일에서 exon, intron 등의 gene 구조를 확인하기 위한 단계

■ Quantification (featureCounts)

Mapping된 bam파일에서 유전자 정보를 담은 GTF annotation files를 통해 각 유전자의 발현량을 수치화 하는 단계